Structure And Biochemical Analysis Of Toxins From The Superbug Clostridium Difficile

Davies, Abigail

Award date:
2014

Awarding institution:
University of Bath

STRUCTURAL AND BIOCHEMICAL ANALYSIS OF TOXINS FROM THE SUPERBUG *CLOSTRIDIUM DIFFICILE*

Submitted by Abigail Davies for the degree of Ph.D.

University of Bath

Department of Biology and Biochemistry

September 2013

Abstract

Clostridium difficile is a gram positive, anaerobic bacterium that is the leading cause of antibiotic-associated pseudomembranous colitis worldwide. *C. difficile* is an extremely infectious bacterium that produces spores that are highly resistant to standard disinfectant agents and can survive on surfaces for long periods of time. Both the resistance of the spores combined with multiple patients with low-immune systems has lead to an increase in hospital-acquired *C. difficile* infection, which has had a severe economic impact on the healthcare system. Due to the emerging antibiotic resistance problems and the common occurrence of patient relapse using the current drugs of choice, alternative therapeutic avenues are being explored. *C.difficile* produces two potent exotoxins; Toxin A and Toxin B that are the causative agents of infection. These toxins have multi-modular domain organisations, with each domain playing a role in cytotoxicity. Some of these domains have been characterised structurally using X-ray crystallography. In this thesis, the low resolution SAXS structure of Toxin A will be presented along with the advances made towards determining the X-ray crystallographic structure of the full-length Toxin A.

In addition to Toxins A and B, some strains of *C. difficile* produce a binary toxin, CDT, which is made up of two individually produced components, CDTa and CDTb. The CDTa component is the enzymatically active component, whereas CDTb is the transport component, directly involved in translocating CDTa into target cells. The precise role of CDT in pathogenesis is unclear, however there is evidence that CDT ADP-ribosylates monomeric actin in target cells, but the detailed mechanism by which this reaction takes place is unknown. Here site directed mutagenesis of key residues of the active site of CDTa was performed and the effect of these mutations on the enzyme's cytotoxicity tested. By separately mutating three active site residues, the cytotoxic effect of CDTa can be completely eradicated, details of which will be discussed in this thesis. Additionally, the progress made towards determining the X-ray crystallographic structure of the transport component, CDTb, will be discussed.

Acknowledgments

I would like to start by thanking both of my supervisors, Professor K. Ravi Acharya and Dr Cliff Shone for giving me the opportunity to work on this project. In particular I'm indebted to Ravi for allowing me to perform the majority of the research in his laboratory and for his support and motivation throughout my Ph.D. I'd also like to thank the BBSRC and PHE for funding.

I greatly appreciate the expertise of Dr Joanna McGlashan and Dr April Roberts from PHE who helped with the mutant Vero cell assays. In addition I'd like to thank Dr Christian Pernstich from the University of Bristol, for donating his time for the SEC-MALS experiments as well as generously lending his expertise for data analysis. I'd like to thank Iain Harris and Ursula Potter at the University of Bath for their help with electron microscopy.

Additionally, I'd also like to acknowledge Dr Jean van den Elsen from the University of Bath, for kindly allowing me to use the Zetasizer Nano S for dynamic light scattering and for his helpful discussions on data analysis. I'd like to thank Harriet Denton and Dr Cliff Shone from PHE for providing generous amounts of Toxin A and Dr Victor Marquez from the Centre for Cancer Research, NCI-Frederick, for kindly providing β-TAD. I would also like to acknowledge Diamond Light Source where X-ray crystallographic data and SAXS data were collected.

I'm thankful to current lab members of 0.34 (Charlotte, Chris, Geoffrey, Hazel, Iain, Liz, Nethaji, Shalini, Sharon, Sneha, Stef, Tram and Will) and past lab members (Ken, Jodie, Akif and Amit) for their support throughout my Ph.D as well as all of the fun memories.

I also appreciate the love and support from my friends at home in Chester, as well as my friends from University for standing by me through the tough times during my Ph.D. I'd especially like to thank my close friends Lois and Justin for making living in Bath an enjoyable and special experience.

I'd like to pay special thanks to Fraser McArthur for his love and support throughout my PhD, including all of the pep talks, motivational speeches, words of encouragement and for his endless belief in me. Although purchased in jest, I still appreciate the 'I was born to do research' T-shirt.

I'm grateful to my family, particularly my grandparents for their endless support and affection throughout my life.

Last but not least, I'd like to thank my incredible parents to whom this thesis is dedicated. Their devotion to helping me achieve my ambitions is never-ending so I'd like to thank them for allowing me to follow my dreams and for their constant support and love provided along the way.

September 2013

Abbreviations

ADPRT	Adenosine Diphosphate Ribosyl Transferase
ARTT	Adenosine Diphosphate Ribosyl Turn-Turn
β-OG	Beta-Octyl-Glucoside
β-TAD	Beta-Methylene-Thiazole-4-carboxamide adenine dinucleotide
CD	Circular Dichroism
CDAD	*Clostridium difficile* Associated Disease
CDI	*Clostridium difficile* Infection
DLS	Dynamic Light Scattering
D_{max}	Maximum Particle Diameter
DO	Dissolved Oxygen
dRI	Refractive Index
GAP	GTPase Activating Protein
GDI	Guanine-Nucleotide Dissociation Inhibitor
GDP	Guanosine Diphosphate
GEF	Guanine-Nucleotide Exchange Factor
GF	Gel Filtration
GTP	Guanosine Triphosphate
HPLC	High Pressure Liquid Chromatography
InsP6	Inositol Hexakisphosphate
LB	Luria-Bertani
LCT	Large Clostridial Toxins
L-CPL	Left-handed Circularly Polarized Light
LS	Light Scattering
LSR	Lipolysis-Stimulated Lipoprotein Receptor
MBP	Maltose-Binding Protein
NSD	Normalized Spacial Discrepancy
OD	Optical Density
PCT	Pre-Crystallisation Test
PDB	Protein Data Bank
PDI	Polydispersity Index
PEG	Polyethylene Glycol
PHE	Public Health England

PID	Proportional-Integral-Derivative
PMC	Pseudomembranous Colitis
P(r)	Radial Patterson Function
PSD	Particle Size Distribution
R-CPL	Right-handed Circularly Polarized Light
R_g	Radius of Gyration
SAXS	Small-Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC-MALS	Size-Exclusion Chromatography Multi-Angle Light Scattering
TB	Terrific Broth
UDP-Glc	Uridine Diphosphate Glucose

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction to *Clostridium difficile*

1.1.1 *Clostridium difficile*

Clostridium difficile Infection

Clostridium difficile is an anaerobic Gram-positive spore-forming bacterium that was first described in 1935 (Hall and O'Toole, 1935). *C. difficile* is named so because it grew slowly in culture and was difficult to isolate (Kelly and LaMont, 1998). It is one of the most common causes of antibiotic-associated diarrhoea in the world (Carroll and Bartlett, 2011; Kyne, 2010). Initially *C. difficile* was not considered to be a particularly harmful pathogen until a rise in the number of cases of PMC (pseudomembranous colitis) in the 1970s. It is known to cause a number of intestinal diseases known as *C. difficile* associated disease (CDAD) or *C. difficile* infections (CDI) (Kyne, 2010). These diseases range from mild diarrhoea to severe toxic megacolon, which can be fatal. Infection with *C. difficile* accounts for 90–100 % of antibiotic-associated PMC, with a 6–30 % mortality rate (McMaster-Baxter and Musher, 2007). CDI is associated with a broad range of symptoms including fever, abdominal pain, and leukocytosis, but patients with CDI are diagnosed by the presence of *C. difficile* toxins in faecal samples. Patients with PMC are treated with broad-spectrum antibiotics such as metronidazole and vancomycin, which can be very effective. Metronidazole and vancomycin have similar efficacy in mild disease, however there is growing concern regarding treatment failure with the use of metronidazole in severe disease where vancomycin has been shown to be superior (Ananthakrishnan, 2011; Aslam *et al.*, 2005; Musher *et al.*, 2005). However metronidazole is commonly the drug of choice as it is a lower cost drug than vancomycin. The use of broad-spectrum antibiotics to treat other diseases can suppress the intestinal bacterial flora causing a disturbance in the normal balance in the gut, resulting in host susceptibility to colonization or overgrowth of *C. difficile* (Bartlett, 2002; Carroll and Bartlett, 2011; Kelly and LaMont, 1998; Lyerly *et al.*, 1988; Shen, 2012; Voth and Ballard, 2005). Patients with mild cases of CDI are usually advised to discontinue taking any antibiotics to prevent suppression of the natural

gut flora, whereas for patients with extreme cases of CDI that lead to PMC, metronidazole and vancomycin are the suggested treatments (Cecil, 2012). Relapse is a common problem associated with antibiotic treatment (occurring in up to 25% of cases), in which case patients are required to continue with prolonged courses of these drugs (Bartlett, 2002). As treatment with antibiotics commonly leads to relapse, an alternative treatment of faecal transplantation is sometimes used whereby fresh donor faeces are introduced into the cecum (Mattila *et al.*, 2012). This helps to re-establish the normal composition of intestinal flora preventing *C. difficile* from colonizing the gut.

C. difficile is an extremely infectious bacterium and can produce highly resistant spores. These spores can survive on surfaces for long periods of times and are resistant to standard disinfectants (Cecil, 2012). Bleach can kill *C. difficile* spores but their resistance and ability to survive for long periods are particularly difficult to control in hospital environments which means the majority of infections are contracted nosocomially (Davies *et al.*, 2011; He *et al.*, 2013). Thus CDIs cause significant problems in the healthcare system, with studies showing that CDI infections cost hospitals in the US an estimated $1-3 billion annually (McGlone *et al.*, 2012).

Clostridium difficile Virulence Factors

C. difficile has a number of virulence factors that contribute to its pathogenicity, such as its flagella, capsule and putative Type IV pilus. In addition *C. difficile* produces a variety of proteins for adhesion and evasion including collagen and fibronectin-binding proteins, von Willebrand Factor binding proteins, sortases, cysteine proteases and adhesins (Vedantam *et al.*, 2012). The pathogenic strains of *C. difficile* produce two potent exotoxins, Toxin A and Toxin B (often called TcdA and TcdB), which induce mucosal inflammation and diarrhoea (Borriello, 1998; Kelly and LaMont, 1998; Voth and Ballard, 2005). In addition to these exotoxins, some of the more virulent *C. difficile* strains produce an ADP- ribosylating binary toxin (CDT), however, the role of this toxin in disease is unclear (Perelle *et al.*, 1997; Popoff *et al.*, 1988). Details of both Toxin A, Toxin B and CDT will be discussed in terms of biology and structural features in Chapter 1.2.

1.2 Introduction to *C. difficile* Toxins

1.2.1 Large Clostridial Toxins A and B

Biology of LCTs

Bacteria from the genus *Clostridium* produce a wide range of toxins, the most well known being the clostridial neurotoxins such as the *Clostridium botulinum* neurotoxins (BoNTs), which can cause flaccid muscular paralysis known as botulism. Additionally there is the well known *Clostridium tetani* neurotoxin (TeNT) which causes spastic paralysis (Pellizzari *et al.*, 1999). *Clostridium difficile* produces two large exotoxins: Toxins A and B, which are part of the large clostridial toxin (LCT) group, owing to their high molecular mass: Toxin A is 308 kDa and Toxin B is 269.6 kDa (Jank and Aktories, 2008). Other key members of the LCT family include, lethal toxins from *Clostridium sordellii* and α-toxin from *Clostridium novyi.* All members of the LCT family are monoglycosyltransferases that inactivate small GTPases in target cells (Busch and Aktories, 2000; Just and Gerhard, 2004).

Toxins A and B are encoded by the genes *tcdA* and *tcdB,* which are located in a 19.6 kb locus, known as the PaLoc (pathogenicity locus) (Aubry *et al.*, 2012; Govind and Dupuy, 2012; Vedantam *et al.*, 2012; Voth and Ballard, 2005). There are two other genes in the PaLoc that encode positive and negative regulators of the production of these toxins, TcdR and TcdC (Matamouros *et al.*, 2007). In addition there is a gene encoding a holin-like protein which is thought to allow secretion of the toxins, TcdE (Govind and Dupuy, 2012; Rupnik *et al.*, 1998; Tan *et al.*, 2001). *C. difficile* strains are grouped on the basis of the variations in the structure of the PaLoc (Rupnik, 2005; Rupnik *et al.*, 1998).

Toxins A and B share 47% sequence identity and have multi-modular domain organisations described by Jank and Aktories as the ABCD model (Jank and Aktories, 2008). The 'A' domain is the biologically active N-terminal domain, the 'B' domain is the C-terminal binding domain, the 'C' domain is the cysteine

protease domain, located next to the 'A' domain, and the 'D' domain is the hydrophobic pore forming domain, located between the 'C' and 'B' domains (for a schematic diagram see Figure 1.2.1a). The C-terminal binding 'B' domain, which consists of polypeptide repeats, is involved in receptor binding to specific cell-surface carbohydrate receptors, but other domains may also play a direct or indirect role in binding (Frisch *et al.*, 2003).

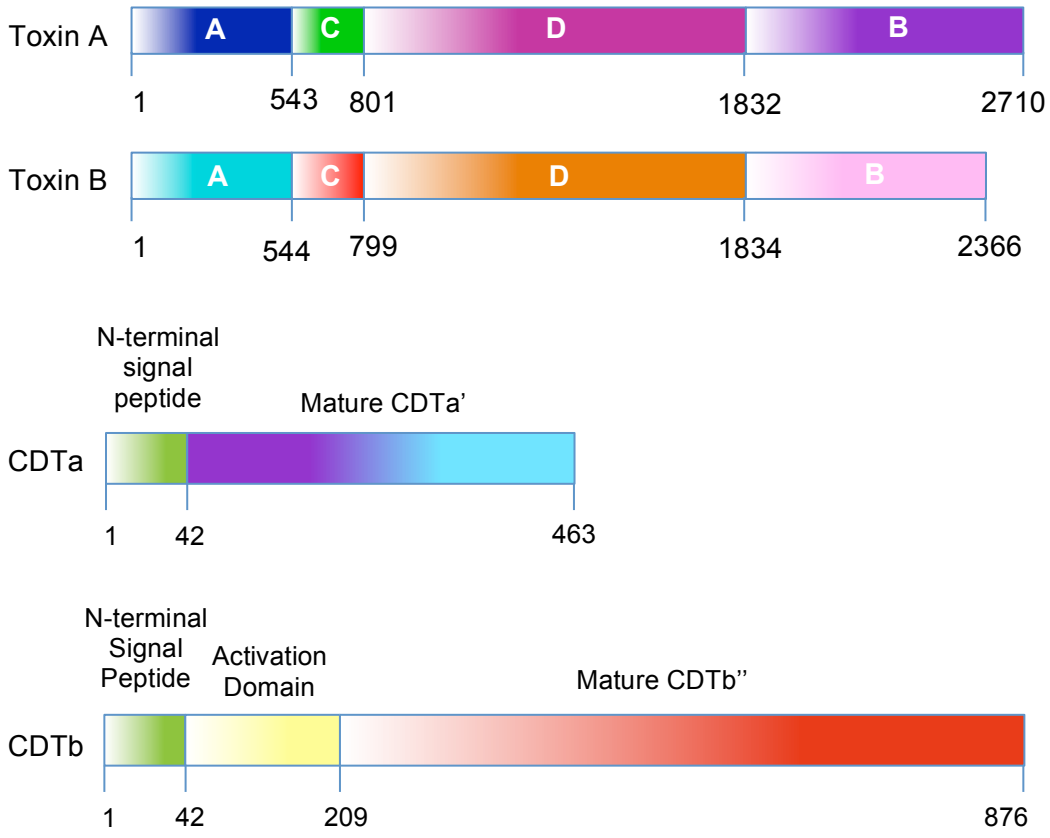


Figure 1.2.1a- Schematic representation of domain organisation of *Clostridium difficile* toxins. A: Domain organisation of Toxin A and Toxin B from the LCT family of toxins. B: Domain organisation of the two components of the binary toxin, CDT.

The precise mechanism by which Toxins A and B enter target cells is unclear, however there is a theory based on multiple findings that can be explained in four key steps, which are summarised in Figure 1.2.1b.

Step 1: The binding 'B' domain interacts with highly specific carbohydrate receptors on the cell surface, which stimulates receptor-mediated endocytosis of the toxin into the endosomal compartment.

Step 2: The pH of the endosome decreases causing conformational changes within the toxin, allowing potential pore formation of the 'D' domain and insertion into the endosomal membrane (Barth *et al.*, 2001; Qa'Dan *et al.*, 2000). The 'A' domain and 'C' domain are translocated into the cytosol which has been shown by the use of ^{86}Rb ions (Barth *et al.*, 2001; Qa'Dan *et al.*, 2000).

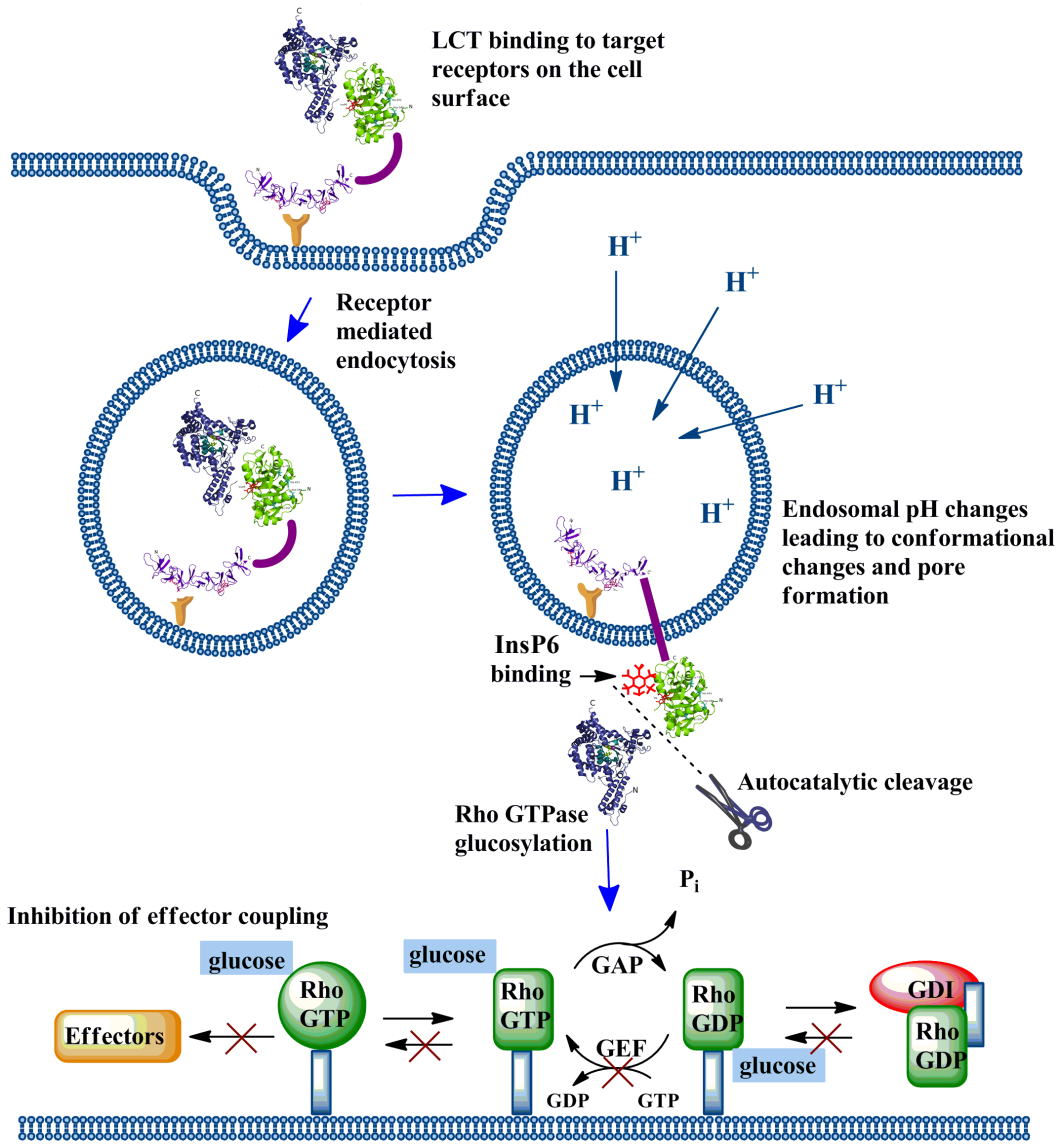


Figure 1.2.1b- Diagram illustrating the predicted mechanism of LCT uptake into target cells including the toxic effect of the 'A' domain (Davies *et al.*, 2011).

Step 3: It is thought that the cysteine protease domain 'C' is involved in autoprocessing of the toxins, and requires InsP6 (Egerer *et al.*, 2007; Giesemann *et al.*, 2008; Pruitt *et al.*, 2009). This autoproteolysis allows the release of the enzymatic 'A' domain into the cytosol.

Step 4: The LCTs target Rho proteins for glucosylation by the biologically active 'A' domain in the cytosol. Toxins A and B specifically target Rho GTPases (Rho, Ras and Cdc42), which are molecular switches involved in numerous signalling processes, in particular the regulation of the actin cytoskeleton (Figure 1.2.1b). The 'A' domains of Toxins A and B catalyse the addition of UDP-Glc (UDP-glucose) to Thr-37 (monoglucosylation) in Rho GTPase leading to depolymerisation of actin filaments, disruption of the cytoskeleton and eventually cell rounding and cell death (Jank *et al.*, 2007a; Just *et al.*, 1995a, 1995b).

There is on-going speculation as to which of the two toxins are essential for virulence, with contradicting reports. There are a number of reports that suggest Toxin B is essential for pathogenicity, whereas the role of Toxin A is less clear. This is reflected in the reported outbreaks of *C. difficile* strains that are Toxin A negative/Toxin B positive. Universal gene knock-out systems developed for the genus *Clostridium* (Heap *et al.*, 2007; O'Connor *et al.*, 2006) were used in studies that indicate Toxin A is not essential for virulence (Lyerly *et al.*, 1985; Lyras *et al.*, 2009). This is also reflected in studies by Kim *et al.*, who reported that there was no significant difference in disease between Toxin A negative/Toxin B positive strains and Toxin A positive/Toxin B positive strains which indicates that Toxin A doesn't have a big impact on pathogenicity (Kim *et al.*, 2012). However, recent reports suggest that both toxins play a role in disease and that Toxin A is able to cause disease in the absence of Toxin B (Kuehne *et al.*, 2011, 2010).

C. difficile is a harmful pathogen and the main cause of the rise in cases of antibiotic-associated PMC in hospitals, therefore it is of great interest to understand the function of LCTs, especially with the upcoming occurrence of antibiotic resistance. Structural studies have delivered an insight into the molecular mechanism of the LCTs, which will be detailed in the following sections.

High-Resolution Structural Aspects of LCTs

X-Ray Crystallography Studies of LCTs

X-ray crystallography is a technique used to determine the structures of molecules at atomic resolution. This experimental technique requires highly ordered crystals of the molecule of choice. Crystals are mounted on a goniometer and are rotated whilst being exposed to a beam of X-rays. The crystals diffract the X-rays producing a series of reflections called the diffraction pattern, which provides information about the arrangement of the molecules within the crystal. The full-length crystal structures of Toxin A and B have not yet been elucidated, however, crystal structures of some of the individual domains have been solved. The successfully solved structures of Toxin A domains are displayed in Figure 1.2.1c and those for Toxin B domains are displayed in Figure 1.2.1d, with a more detailed analysis of structures provided in the following sections.

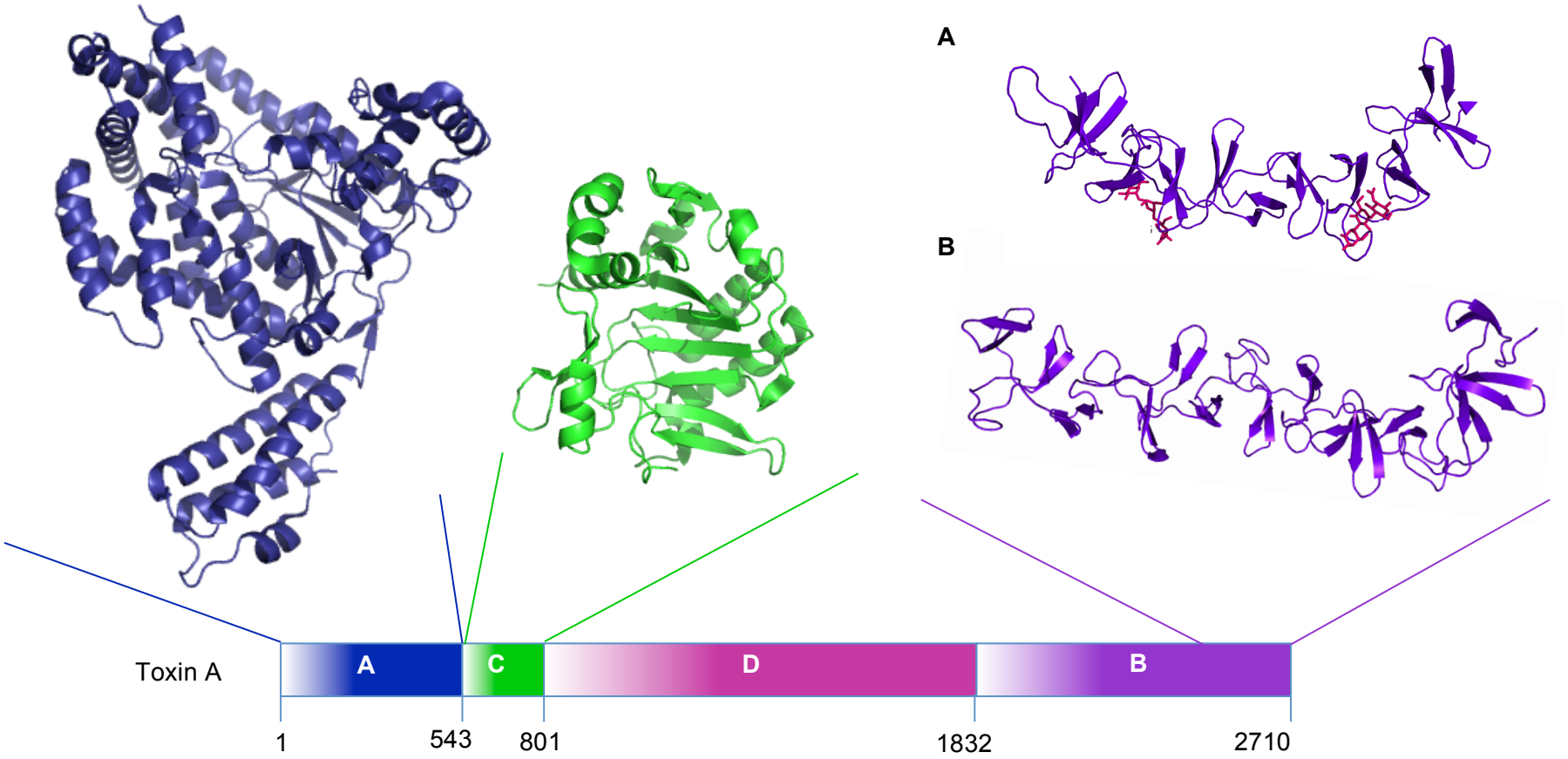


Figure 1.2.1c- Crystal structures of the 'A' domain, 'C' domain and 'B' domain of Toxin A. **Structures of the individual domains are coloured corresponding to the schematic diagram of domain organisation. PDB codes: 4DMW and 3HO6 are 'A' domain and 'C' domain respectively. A: A smaller version of the 'B' domain with CD-grease (synthetic derivative of a naturally occuring carbohydrate receptor) bound (PDB code: 2G7C). B: Longer version of 'B' domain (PDB code: 2QJ6)(D'Urzo *et al.*, 2012; Greco *et al.*, 2006; Pruitt *et al.*, 2009). Images were created using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

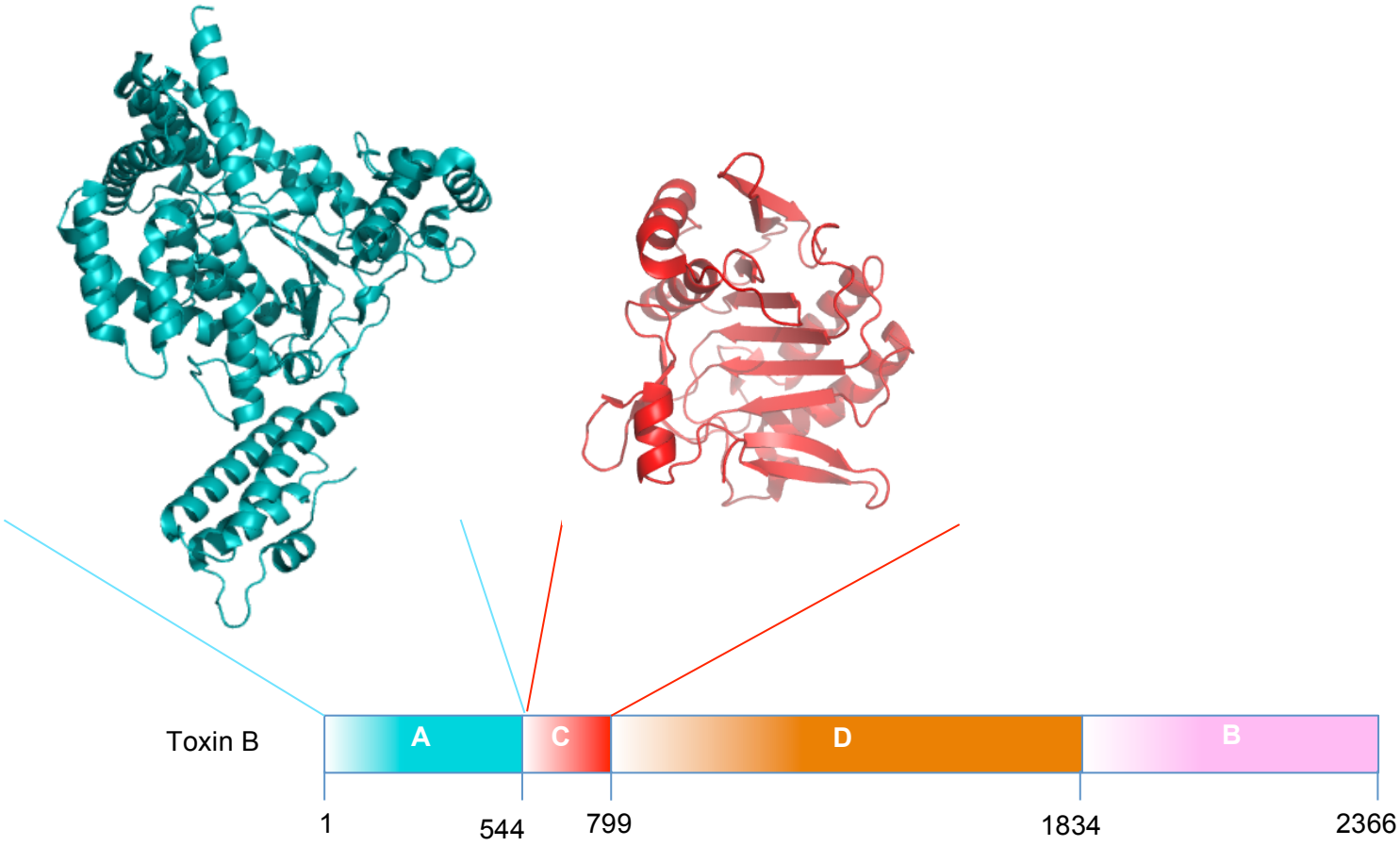


Figure 1.2.1d- Crystal structures of the 'A' domain and 'C' domain of Toxin B. **Structures are colour coordinated with the schematic representation of domain organisation. PDB codes: 2BVL and 3PEE representing the 'A" domain and 'C' domain respectively (Reinert *et al.*, 2005; Shen *et al.*, 2011). Images were created using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

Glucosyltransferase 'A' Domain

The glucosyltransferase domain, which is responsible for glucosylation of Rho proteins, is located at the N-terminal region known as the 'A' domain. The crystal structures of both the 'A' domain of Toxin A and the 'A' domain of Toxin B have been solved with resultant structures displayed in Figures 1.2.1c and 1.2.1d respectively, PDB codes 4DMW and 2BVL (D'Urzo *et al.*, 2012; Reinert *et al.*, 2005). This catalytic fragment of the LCTs is delivered to the cytosol where it glucosylates small GTPases.

The 'A' domain of Toxin A consists of residues 2-538 and was crystallised in the presence of UDP-Glc and manganese ions, however density was not observed for either the UDP-Glc or the manganese ions (D'Urzo *et al.*, 2012). The crystal structure shares the same common fold (consisting of 234 residues) seen among the glucosyltransferase type A family. The remaining residues are arranged predominantly in α-helices and are thought to contribute to the specificity of the toxin.

The 'A' domain of Toxin B consists of residues 1–543 and was co-crystallized with UDP-Glc and manganese ions that are essential for catalysis (Ciesla and Bobak, 1998; Rupnik *et al.*, 2005). The hydrolysed products of UDP-Glc; α-D-glucose and UDP were identified in the electron density due to hydrolytic activity of the 'A' domain (Reinert *et al.*, 2005).

These two crystal structures of the 'A' domains share almost identical overall folds with a root mean square deviation (rmsd) value of 1 Å (D'Urzo *et al.*, 2012). Therefore as the 'A' domain of Toxin B was solved with the hydrolysis product of UDP-Glc, the structure and active site will be looked at in more detail. The active site of Toxin B is displayed in Figure 1.2.1e as a stereo image, with the residues displayed as sticks. The catalytic pocket is made up of the β2, β5 and β10 strands, in addition to α12 and α18 helices, and a 510–523 residue loop (Ünligil and Rini, 2000). The catalytic DXD-motif involved in the binding of UDP-Glc and Mn^{2+} is composed of Asp-286 and Asp-288. Asp-286 forms a hydrogen bond with a water molecule of the Mn^{2+} co-ordination sphere, in addition to binding to both the 3′-hydroxy group of UDP-ribose and glucose, making it a key residue for catalysis, whereas the Asp-288 residue binds directly to Mn^{2+}

(Reinert *et al.*, 2005). Mutational analysis by alanine scanning revealed five important residues specifically involved in enzyme activity: Asp-270, Arg-273, Tyr-284, Asn-384 and Trp-520. In addition, the following four residues were reported to be important for UDP-Glc recognition: Arg-455, Asp-461, Lys-463 and Glu-472 (Jank et al., 2005).

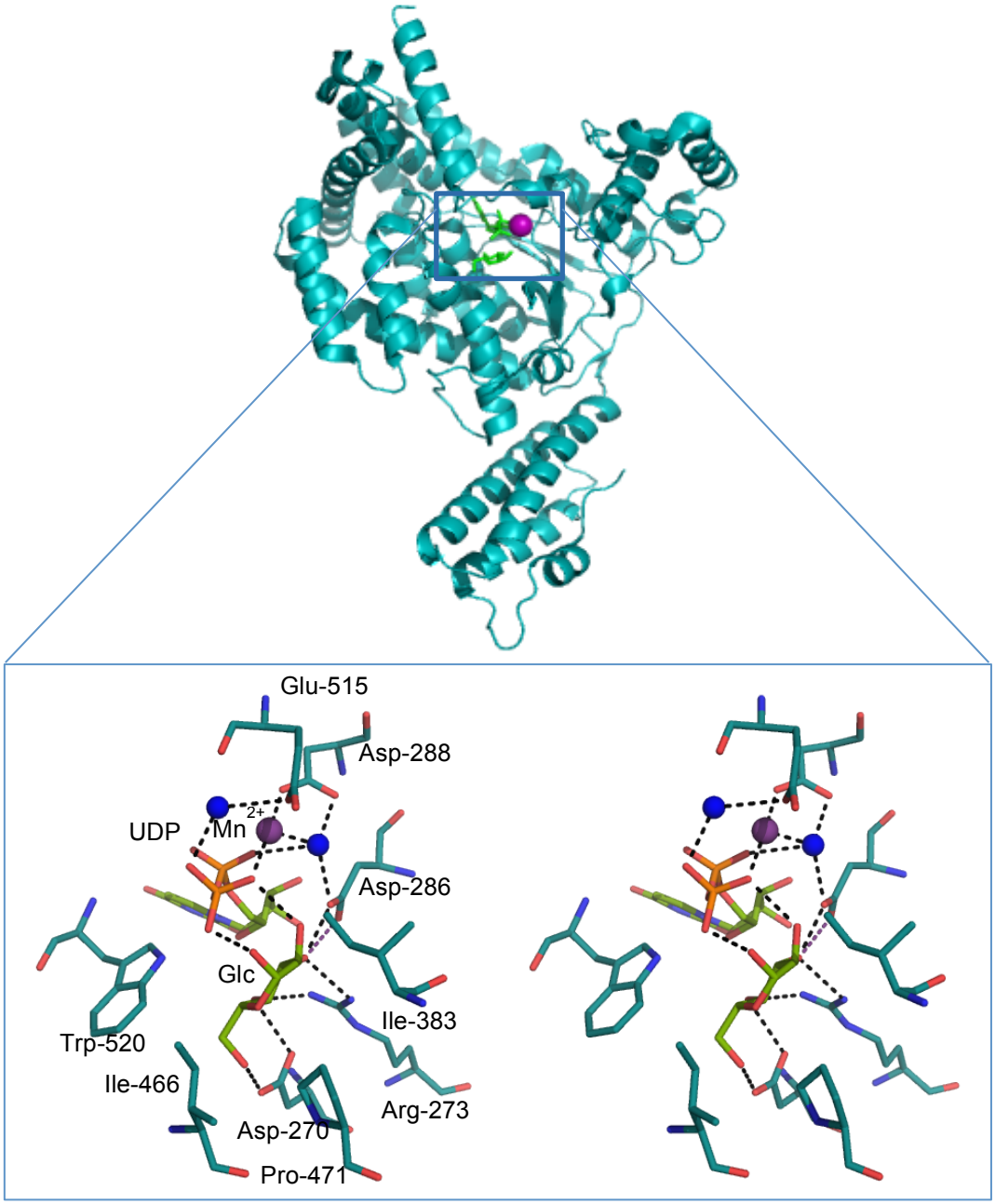


Figure 1.2.1e- Crystal structure of the 'A' domain of Toxin B. A: Structure of Toxin B 'A' domain displayed in cyan with α-D-glucose and UDP displayed as green sticks, and the manganese ion as a magenta sphere. B: A zoomed in stereo view of the catalytic pocket with the α-D-glucose as green sticks and UDP displayed as orange sticks, the coordinating water molecules as blue spheres and the manganese ion as a magenta sphere (Reinert *et al.*, 2005). **Images were created using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

Toxins A and B of the LCT family target and glucosylate the Rho family of GTPase proteins. These proteins are molecular switches that are predominantly responsible for regulating the actin cytoskeleton, but also play roles in cell polarity, gene transcription and cell cycle progression. The Rho GTPase family of proteins consists of the Rho, Rac and Cdc42 subfamilies. They can exist in two forms: firstly the GTP-bound form which is active and secondly the GDP-bound form which is inactive, as displayed in Figure 1.2.1b. The GTP and GDP-bound forms are inter-convertible and are regulated by guanine-nucleotide-exchange factors (GEFs), guanine-nucleotide-dissociation inhibitors (GDIs) and GTPase-activating proteins (GAPs) (Figure 1.2.1b). GDIs bind to Rho GTPases at a switch I region, rendering them soluble in the cytosol and in the inactive form. GEFs activate GTPases to their GTP-bound form allowing interaction with numerous effector molecules that control a number of signalling pathways. GAPs hydrolyse GTP-bound Rho GTPases into inactive GDP-bound Rho GTPases (Bishop and Hall, 2000). LCTs target the Thr-37 residue on Rho GTPases, which is exposed when in the GDP-bound conformation. This residue is glucosylated by the 'A" domain of the toxins, which causes depolarization of actin filaments, leading to disruption of the cytoskeleton and cell death. The Thr-37 residue is not exposed when in the GTP-bound form or when their switch I region is complexed with GDIs. To confirm, Rho is glucosylated in the GDP-bound conformation by Toxins A and B, as the Thr-37 residue is exposed which allows an oxygen acceptor atom to be in position for attack from the C1 donor atom of the LCTs (Just *et al.*, 1995a, 1995b). Although there are reports that indicate that Toxin B can bind to RhoA in the Rho-GTP form (Jank *et al.*, 2007b).

There are two types of mechanism proposed for the glucosylation mechanism of the 'A' domain: Firstly, a double displacement reaction (SN2-like) and secondly an internal return stereo-specific (SN1) reaction. It is unlikely that the glucosylation occurs via an SN2-like mechanism, as this requires a nucleophile in position to attack the β-side of the C1 atom in glucose. However, as this atom is surrounded by non-polar residues in the catalytic cleft, there is no nucleophile in position to attack the C1 atom, hence a double displacement reaction is unlikely (Davies *et al.*, 2011). An SN1 is plausible as the oxocarbenium intermediate could be stabilised by the surrounding negative phosphate and carboxylate groups from Asp-270, Asp-286, Asp-288 and Glu-515, which are in turn compensated by the positively charged Arg-273 and the Mn^{2}+ ion (Reinert *et al.*, 2005). LCTs exhibit high substrate/co-substrate

specificity for UDP-Glc, with two residues Ile-383 and Glu-385, located in the active site being responsible for this specificity as shown by mutagenesis studies (Jank *et al.*, 2007b, 2005).

Binding 'B' Domain

The initial step in the toxication process is the binding of LCTs to specific cell-surface receptors of intestinal epithelial cells, which is carried out by the 'B' domain. This 'B' domain interacts with carbohydrate structures such as Gal-α-(1,3)-Gal-β-(1,4)-GlcNAc on the host epithelial cells (Krivan *et al.*, 1986). Although a functional carbohydrate receptor in humans has yet to be identified, attempts to gain more insights into carbohydrate binding have been made (Davies *et al.*, 2011). The crystal structure of the C-terminal binding domain of Toxin A has been determined for two different fragment sizes. The first fragment (127 residues), named TcdA-f1, was obtained by expressing a longer fragment and isolating a smaller cleaved fragment (PDB code 2F6E) (Ho *et al.*, 2005). The second fragment, TcdA-f2, which is a slightly longer fragment (255 residues), was solved in complex with a synthetic carbohydrate derivative, CD-grease [α-Gal-(1,3)-β-Gal-(1,4)-β-GlcNAcO(CH)COCH] (Figure 1.2.1f) (PDB code: 2G7C) (Greco *et al.*, 2006).

In the TcdA-f2-CD-grease complex there are two carbohydrate-binding regions, however there are seven of these highly conserved binding regions in the full-length C-terminal fragment, which gives the 'B' domain its high binding capacity (Greco *et al.*, 2006; Ho *et al.*, 2005).

Although there is little information about the binding domain of Toxin B, it is believed that Toxin B uses different receptors to bind to target cell surfaces than Toxin A, which may explain the differences observed in toxicity as discussed earlier (Jank *et al.*, 2007b). However, as Toxins A and B share 47% sequence identity, the structures of the binding domains may share similarities and therefore predictions can be made that Toxin B will interact with carbohydrate receptors in a similar manner to Toxin A (Davies *et al.*, 2011).

Sequence analysis of the Toxin A gene shows that 31.5% of the gene is in 38 contiguous repeating units and these repeats are located in the 'B" domain region. These repeats are split into two categories: firstly, class I (long repeats),

of which there are seven types, each repeat being 30 amino acids long; and secondly, class II (short repeats), of which there are thirty-one types, each consisting of either 20 or 21 amino acids (Dove *et al.*, 1990). TcdA-f1 consists of four short repeats and one long repeat whereas TcdA-f2 contains nine small repeats and two long repeats (Greco *et al.*, 2006; Ho *et al.*, 2005). Each short and long repeat form β-hairpins that are connected by loops, see Figure 1.2.1f. In the short repeat regions the loops are made up of 7–10 residues, whereas in the long repeat regions the loops are made up of 18 residues (Jank *et al.*, 2007b).

The β-hairpin repeat regions are related to each other by a screw-axis transformation, meaning that each β-hairpin is related to the adjacent hairpin by 120 ° (Figure 1.2.1f) (Greco *et al.*, 2006). This results in a β-solenoid left-handed helix, which is a common conformation of many bacterial cell-surface-binding proteins. The helical conformation increases the surface area of the 'B' domain optimised for binding to target cells (Figure 1.2.1f) (Greco *et al.*, 2006). These results are consistent with that of the short fragment TcdAC26–39 structure of the 'B' domain from Albesa-Jové´ *et al.* (Albesa-Jové *et al.*, 2010).

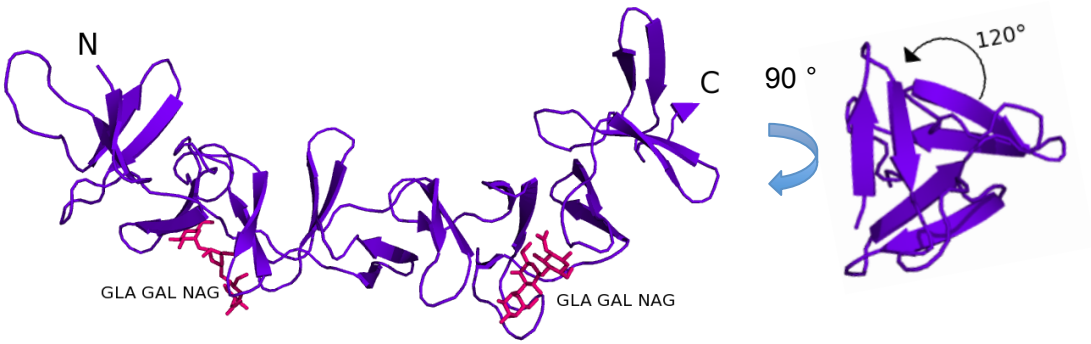


Figure 1.2.1f- Crystal structure of the binding 'B' domain of Toxin A (PDB code: 2G7C). The binding domain is displayed as purple cartoon and the bound CD-grease molecule as magenta sticks. The structure has a left-handed helix shape increasing surface area for binding. The cross section through the helix is displayed illustrating the 120 ° rotation between the β-hairpins creating the helix. In this structure there are two long repeat regions (class I) and nine short repeat regions (class II), which gives a total of eleven β-hairpins. CD-grease molecules are bound in the shallow troughs made up of the long repeat regions and the hairpin turn of the following short repeat region (Greco *et al.*, 2006)**. Images were created using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

Cysteine Protease 'C' Domain

The cysteine protease 'C' domain (CPD) is thought to be located between residues 543 and 769 in Toxin A and between residues 543 and 767 in Toxin B (Egerer *et al.*, 2007). The crystal structure of the 'C' domain (543–809) from Toxin A was determined at 1.6 Å in the presence of inositol-hexakisphosphate (InsP6) (Figure 1.2.1g) (PDB code: 3HO6) (Pruitt *et al.*, 2009). It is postulated that the 'C" domain is involved in autocatalytic cleavage of Toxins A and B in the presence of InsP6. This autoproteolysis releases the 'A' domain into the cytosol where it monoglucosylates Rho-GTPases (Kreimeyer *et al.*, 2011).

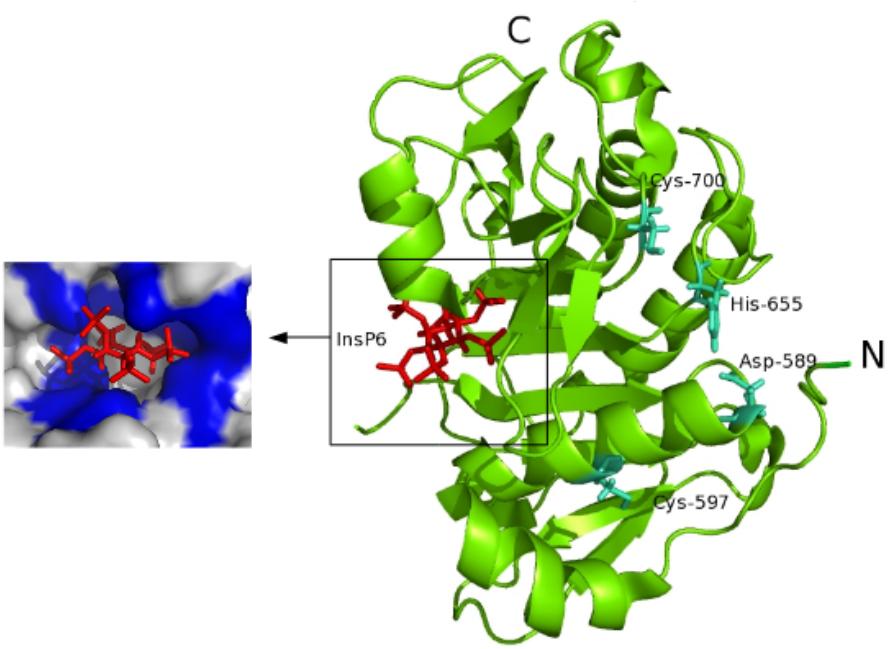


Figure 1.2.1g- Crystal structure of the 'C' domain of Toxin A with InsP6 bound. The 'C' domain is displayed as green cartoon with InsP6 displayed as red sticks. The proposed active site residues are displayed as cyan sticks. The InsP6 binding pocket is also displayed as surface with positively charged residues displayed in blue (Pruitt *et al.*, 2009)**. Images were created using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

The overall structure reveals a central nine-stranded β-sheet flanked by five α-helices. The InsP6-binding site, displayed in Figure 1.2.1g is located on the opposite side of the central β-sheet to the active site. These two binding sites are separated by a three-stranded β-hairpin, known as the 'β-flap'. The N-terminal domain extends around the exterior of the domain from the InsP6-

binding site to the proposed active site (Pruitt *et al.*, 2009). The characteristic cysteine protease catalytic triad has been identified as Cys-700, His-655 and Asp-589; however, these residues are too far apart for hydrogen bond formation, which is normally observed amongst cysteine proteases. Therefore the mechanism of catalysis may differ from other cysteine proteases (Davies *et al.*, 2011). Pruitt *et al*. have shown that the 'C' domain shows significant changes in conformation upon InsP6 binding to a more stable form, which increases resistance to chymotrypsin digestion (Pruitt *et al.*, 2009). InsP6 is negatively charged and binds to a number of positively charged residues that span the entire domain including Arg-753, Tyr-579 and seven lysine residues: Lys-577, Lys-602, Lys-649, Lys-754, Lys-766, Lys-777 and Lys-794 (see Figure 1.2.1g). The 'β-flap', described earlier, is thought to be involved in relaying structural changes from the InsP6-binding site across to the active site, which activates an intramolecular autoproteolytic cleavage event. This allows correct processing of the toxins to release the enzymatic domain into the cytosol (Pruitt *et al.*, 2009).

Figure 1.2.1h- 'C' domain of Toxin A and 'C' domain of Toxin B. Superposition of Toxin B 'C' domain (PDB code: 3PEE) displayed in red and Toxin A 'C' domain (PDB code: 3HO6) displayed in green, with rmsd 1.3 Å (Pruitt *et al.*, 2009; Shen *et al.*, 2011)**. Images were created using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

Shen *et al*. recently solved the structure of the 'C' domain of Toxin B (PDB code: 3PEE) which shares 56% sequence identity with the 'C' domain of Toxin A and a highly similar structure [rmsd (root mean square deviation) of 1.3 Å] (Shen *et al.*, 2011). The two structures share a conserved active site, a well-conserved β-flap and highly similar InsP6 binding sites. Using chemical probes, Shen *et al*. showed that the 'C' domain of Toxin B can adopt the activated conformation without the presence of InsP6, and in fact the presence of InsP6 shifts the equilibrium in favour of the active conformation. Shen *et al*. have used tryptophan fluorescence assays, structural analysis and mutagenesis studies to show that a conserved group of residues in the β- flap region are responsible for transmitting the InsP6-binding signal to the active site (Shen *et al.*, 2011).

Hydrophobic Pore-Forming 'D' Domain

There is little known about the 'D' domain other than that it harbours the majority of the toxin's hydrophobic residues. It is thought to be involved in pore-formation in the endosomal membrane allowing translocation of part of the LCTs into the cytosol. The 'D' domain has not been characterised structurally and there is no structural data for homologous proteins either. Owing to the hydrophobic nature of the 'D' domain, purification and crystallization are difficult, which may explain why the structure has not yet been determined (Davies *et al.*, 2011). Qa'Dan *et al*., used bafilomycin A1, which is a potent inhibitor of the endosomal vacuolar ATPase pump that controls the acidity of the endosome, to assess the effect of endosomal acidification on the LCTs (Qa'Dan *et al.*, 2000). Translocation of LCTs was measured using TNS [2-(p-toluidinyl)napthalene-6-sulfonic acid, sodium salt] fluorescence. For Toxin B, Qa'Dan *et al*., reported that increased TNS fluorescence was observed upon increased acidity in the endosome, suggesting a conformational change within Toxin B and exposure of the hydrophobic 'D' domain, thereby confirming its role in translocation (Qa'Dan *et al.*, 2000).

Low-Resolution Structural Aspects of LCTs

Small-Angle X-Ray Scattering

Due to the lack of full-length X-ray diffraction data for the LCTs, small-angle X-ray scattering (SAXS) has been employed to provide low-resolution structures. Recently reported SAXS data for Toxin B have provided an *ab initio,* low resolution surface model, see Figure 2.2.1d (Albesa-Jové *et al.*, 2010). The four individual domains described in the ABCD model have been visualised for the first time within the SAXS envelope. The data revealed four distinct domain boundaries, into which the crystal structures can be aligned. The four domains are organised as demonstrated in Figure 1.2.1a. The SAXS envelope is elongated in structure and has a protruding region thought to be that of the hydrophobic 'D' domain (Albesa-Jové *et al.*, 2010). The SAXS data were collected at a neutral pH, so the low pH-induced conformational change that is predicted to occur within this 'D' domain has not been shown.

Negative Stain Electron Microscopy

Recent studies have shown that Toxins A and B have similar native structures as determined by electron microscopy (Pruitt *et al.*, 2010). Both toxins have an elongated structure containing a 'head' domain, a long 'tail' domain and a short inner 'tail' domain. Mapping studies were focused on Toxin A as this protein was considerably more homogenous in structure than Toxin B. The three-dimensional structure of Toxin A was constructed using the random conical tilt approach where the modular domains were mapped into the corresponding structure (Pruitt *et al.*, 2010). Pruitt *et al*. concluded that the long 'tail' fragment corresponds to the 'B' domain, the short 'tail' corresponds to the 'A' domain and the 'head' corresponds to the 'D' domain using antibody labelling (Pruitt *et al.*, 2010). The cysteine protease 'C' domain could not be located. Unlike the SAXS structure, electron micrographs were taken of the toxins at a reduced pH and as predicted, conformational changes were observed in the 'head' shape, confirming the location of the 'D' domain in the structure (Pruitt *et al.*, 2010). Although negative stain electron microscopy provides images of the overall toxin shape, it is beneficial to understand the toxins at a molecular level using X-ray crystallography.

1.2.2 Binary Toxin

Biology of Binary Toxin

In addition to the LCTs, some *C. difficile* strains can produce an additional toxin called the ADP-ribosyltransferase binary toxin, CDT (Popoff *et al.*, 1988). CDT is a member of the ADPRT family of binary toxins which has been divided into four key classes: the AB_5 group, the AB-3 group, the single-polypeptide group and the AB group (Holbourn *et al.*, 2006). The AB_5 group that includes cholera toxin, are composed of one A subunit and five B subunits, and they target small regulatory G-proteins. The AB-3 group that includes diphtheria toxin have a binding domain, a transmembrane domain and a catalytic domain, and they ribosylate a diphthamide residue on elongation factor 2. The single polypeptide group, which includes C3 toxin from *C. botulinum* are not well characterised and have unknown roles in pathogenesis. The AB group, which includes C2 toxin from *C. botulinum* and Iota toxin from *Clostridium perfringens,* are made up of two subunits and target actin (Holbourn *et al.*, 2006). CDT is a member of the AB binary toxin group and is made up of two independently produced components: the enzymatic component, CDTa (mature length 48kDa), and the transport component, CDTb (mature length 74kDa); see Figure 1.2.1a for a schematic representation of the domain organization (Barth, 2004; Davies *et al.*, 2011; Perelle *et al.*, 1997)

The precise role of the CDT binary toxin in pathogenesis is unclear. Most research has focused on the LCTs which are the dominant exotoxins produced by *C.difficile* and are known to play key roles in pathogenesis. In general the most virulent *C. difficile* strains maintain the dominant production of LCTs and lack of CDT (Perelle *et al.*, 1997). However, in one study of CDT genes in the United States, 15.5% of *C. difficile* strains tested had both the genes for the CDT components (*cdtB* and *cdtA),* 8.7% of which did not have the LCT genes (Geric *et al.*, 2003). Therefore, the binary toxin may be playing a role in pathogenesis. It has been shown that a CDTa–CDTb (CDT) complex is toxic to Vero cells (Sundriyal *et al.*, 2010). Using time-lapse and immunofluorescence microscopy, Schwan *et al.,* have shown that CDT forms dynamic microtubule protrusions on the surface of human colon carcinoma cells (Caco-2) concomitantly with ADP-ribosylation of actin and depolymerization of microfilaments (Schwan *et al.*,

2009). It is thought that the binary toxin increases adherence of the bacteria to the intestinal epithelial cells using these cell surface extensions. Schwan *et al.*, noted that the protrusions form a dense meshwork in which the bacteria were caught, contributing to the colonization of *C. difficile.* Similar results were also demonstrated for the homologues *C. botulinum* toxin C2 and *C. perfringens* Iota toxin (Schwan *et al.*, 2009). In addition to protrusion formation, cellular microtubule structures were also altered to increase bundling of microtubules. The two components of CDT play different roles in the toxicity process: the transport component, CDTb, is responsible for transport of the CDTa component into the cytosol, whereas the CDTa (enzymatic component) is responsible for ADP-ribosylation of actin in the cytosol (Barth *et al.*, 2004). The proposed mechanism of uptake of CDTa is similar to other members of the ADPRT family such as *C. botulinum* C2 toxin and Iota toxin from *C. perfringens* (Barth, 2004). The CDTb component must be activated via cleavage, after which the proposed mechanism is as follows:

Step 1: The activated mature CDTb'' forms heptamers at the cell surface and binds to specific cell-surface receptors, see Figure 1.2.2a.

Step 2: CDTa binds to CDTb and together they are taken up into the cell by receptor-mediated endocytosis.

Step 3: The pH of the endosomal compartment decreases causing conformational changes that allow pore formation and insertion into the endosomal membrane.

Step 4: CDTa is translocated through the pore into the cytosol where it I rreversibly ADP-ribosylates monomeric G-actin (Barth, 2004; Barth *et al.*, 2004).

The N-terminus of CDTa is responsible for interaction with CDTb, whereas the C-terminus harbours the enzymatic activity (Gulke *et al.*, 2001). Structural evidence from the complex of actin with the enzymatic component of the Iota binary toxin (Ia) shows that Arg-177 of actin is the ribosylation site (Tsuge *et al.*, 2008). Therefore as Ia and CDTa are close homologs, we can predict that CDTa will irreversibly ADP-ribosylate monomeric G-actin at the Arg-177 residue. This ADP-ribosylation blocks polymerisation of G-actin to F-actin and subsequently disrupts the F-actin:G-actin equilibrium (Mauss *et al.*, 1990; Vandekerckhove *et al.*, 1987). Disruption of the actin equilibrium results in collapse of the cytoskeleton, cell rounding and cell death.

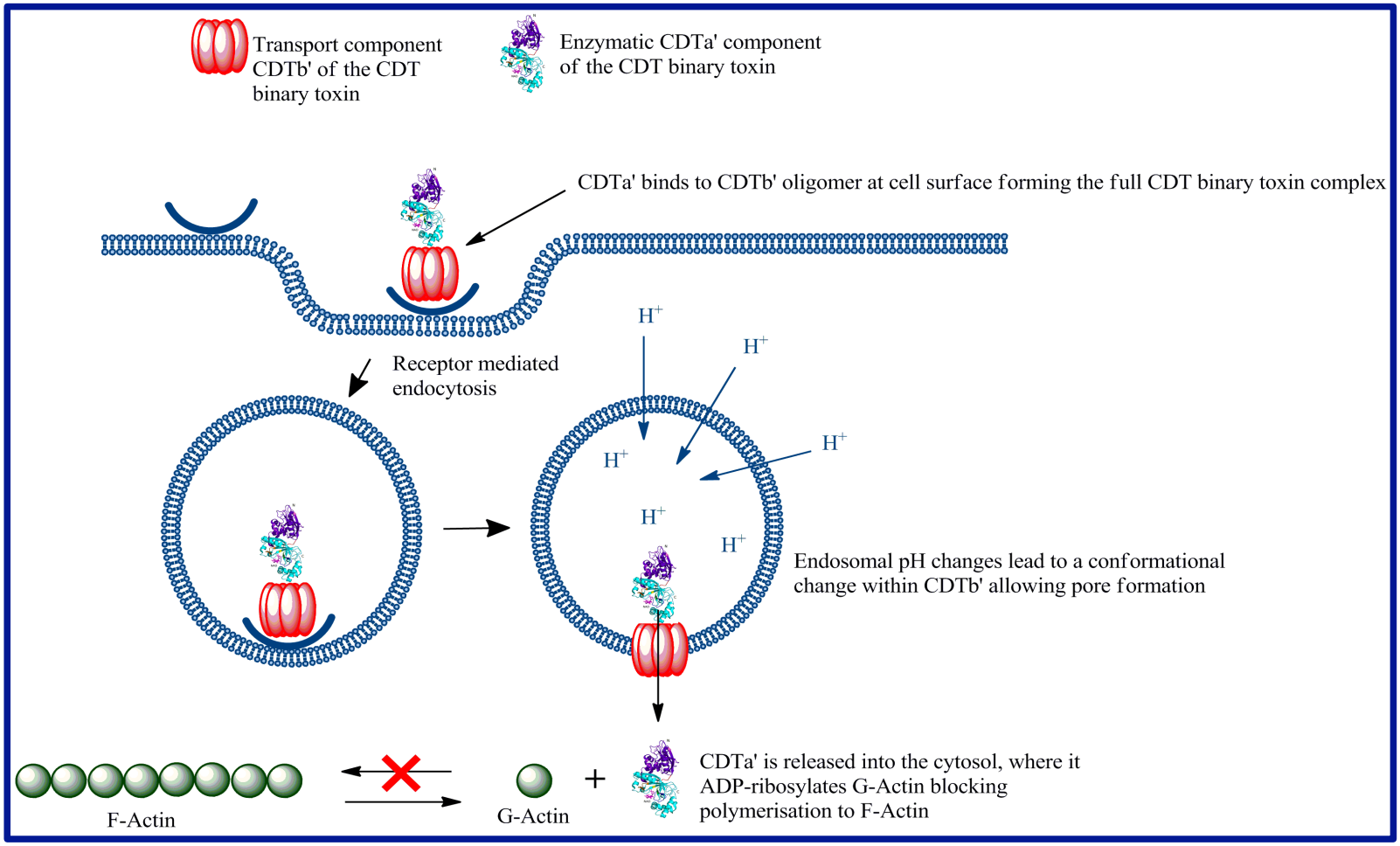


Figure 1.2.2a- The proposed mechanism of entry of CDT into target cells. **The trypsin activated CDTb'' transport component forms heptamers at the target cell surface binding to target cell receptors. CDTa binds to the transport component and together they are taken up into endosomes by receptor-mediated endocytosis. Upon decreasing pH in the endosomal compartment, the CDTb'' heptamer undergoes conformational changes forming a pore which allows translocation of CDTa into the cytosol where it mono ADP-ribosylated G-actin.**

Structural Aspects of Binary Toxin

The crystal structure of CDTa was solved at three different pH values, 4.0, 8.5 and 9.0 (PDB codes 2WN8, 2WN4 and 2WN5 respectively), and in complex with NADPH and NAD at pH 9.0 (PDB codes: 2WN6 and 2WN7) (Sundriyal *et al.*, 2009). The structure of CDTa in complex with NAD, displayed in Figure 1.2.2b, was determined at 2.25 Å. Both the structure of CDTa combined with the mechanism of ADP-ribosylation of *C. perfringens* Iota toxin (with which CDTa shares 84 % sequence identity), have been used to propose a detailed mechanism of ADP-ribosylation by CDTa (Mauss *et al.*, 1990; Sundriyal *et al.*, 2009; Vandekerckhove *et al.*, 1987). It is thought that CDTa transfers the ADP ribose group of NAD/NADPH to monomeric G-actin at Arg-177, blocking polymerization of actin and therefore leading to the collapse of the cell cytoskeleton (Davies *et al.*, 2011).

The structure of CDTa in complex with NAD is displayed in Figure 1.2.2b. The N-terminal domain, which consists of five α-helices and eight β-strands, extending from residues 1 to 215, is displayed in purple and is thought to interact with CDTb. The C-terminal domain, which extends from residues 224 to 240, also consisting of five α-helices and eight β-strands, is displayed in cyan and is thought to interact with actin. These domains are linked by a loop extending from residues 216 to 223, shown in red in Figure 1.2.2b (Sundriyal *et al.*, 2009). Both NAD and NADPH bind to the catalytic cleft of CDTa via the interacting residues Arg-302, Arg-303, Gln-307, Asn-342 and Ser-345 (Sundriyal *et al.*, 2009). Figure 1.2.2b displays these catalytic residues in addition to a number of other defining features of the ADPRT family of toxins, for example the PN-loop (orange) and the Arg-motif (yellow). The ExE-motif which is thought to be the key motif involved in the ADP-ribosylation mechanism, is located on the highly flexible ARTT-loop, which is displayed in blue, see Figure 1.2.2b. The ARTT-loop is an ADP-ribosyl turn-turn loop that is important for substrate binding. In CDTa, the EXE-motif is composed of the residues Glu-385 and Glu-387, which are thought to be involved in stabilising the substrate–enzyme complex as confirmed with the corresponding residues Glu-378 and Glu-380 in Iota toxin (Tsuge *et al.*, 2008). The unusual arrangement in this structure is that these residues are not in direct contact with NAD or NADPH, which might suggest that the EXE-motif in CDTa is not necessary for ligand binding and stabilization of this complex, although there is

no experimental evidence to support this. However, this could be due to the location of both Glu-385 and Glu-387 on the flexible ARTT-loop (Davies *et al.*, 2011).

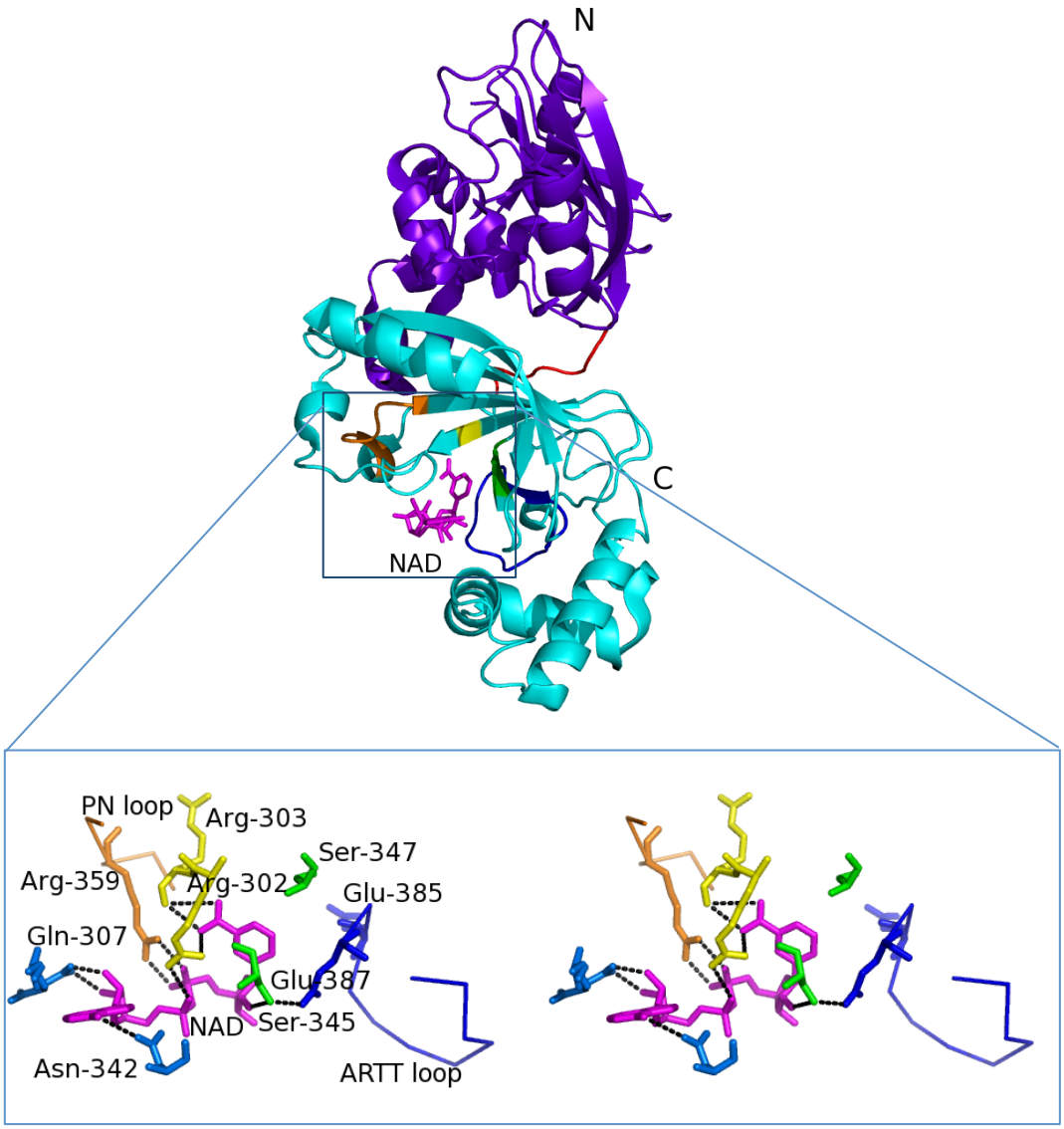


Figure 1.2.2b- Crystal structure of CDTa, the enzymatic component of the C. difficile binary toxin, CDT. The N-terminal domain is displayed in cartoon purple and the C-terminal domain is displayed in cartoon cyan. The active site is zoomed in and displayed in stereo, with the key active site structural motifs colour coordinated in sticks as follows: PN-loop in orange, Arg-motif in yellow, ARTT-loop and ExE in dark blue, STS-motif in green and remaining residues in blue. NAD is displayed as magenta sticks (Sundriyal *et al.*, 2009)**. Images were created using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

The final feature shown in Figure 1.2.2b is the STS-motif, which includes the residues Ser-345 and Ser-347. Ser-345 forms a strong hydrogen bond with Glu-387, and also directly with NAD and NADPH, therefore suggesting a role for the STS-motif in ligand binding and catalysis (Davies *et al.*, 2011; Sundriyal *et al.*, 2009).

An SN1 mechanism has been proposed for ADP-ribosylation of monomeric G-actin. The mechanism was based on the mutagenesis studies of Iota Ia toxin and the crystallographic structure of actin in complex with Ia toxin (Tsuge *et al.*, 2008). The SN1 reaction occurs via two intermediates: firstly an oxocarbenium ion intermediate and secondly a cationic intermediate. This mechanism is thought to be consistent amongst the ADPRT family of toxins (Holbourn *et al.*, 2006). As Iota toxin is the closest homologue to CDT, a similar SN1 reaction has been proposed with the EXE-motif playing a key role (Sundriyal *et al.*, 2009).

The proposed mechanism is as follows (Figure 1.2.2c):

Step 1: The NAD molecule enters the active site of CDTa where the Glu-387 residue from the ExE-motif is positioned near the ribose group of NAD.

Step 2: The catalytic glutamate Glu-387 forms a H-bond with the 2′-OH group of ribose, which leads to loss of the nicotinamide group.

Step 3: This leads to the formation of an oxocarbenium intermediate rendering the ribose group vulnerable to nucleophilic attack, see Figure 1.2.2c. Following cleavage of NAD and formation of an oxocarbenium ion, the ARTT-loop is rearranged to bring the Glu-385 residue to the reaction centre. Arg-177 from ac tin is brought into the active site by Glu-385.

Step 4: The Glu-385 residue stabilises the transfer of ADP-ribose to Arg-177 of actin.

Further structural and experimental evidence is required to validate this hypothesis, such as site-directed mutagenesis of catalytic residues in addition to solving the structure of a CDTa–actin complex.

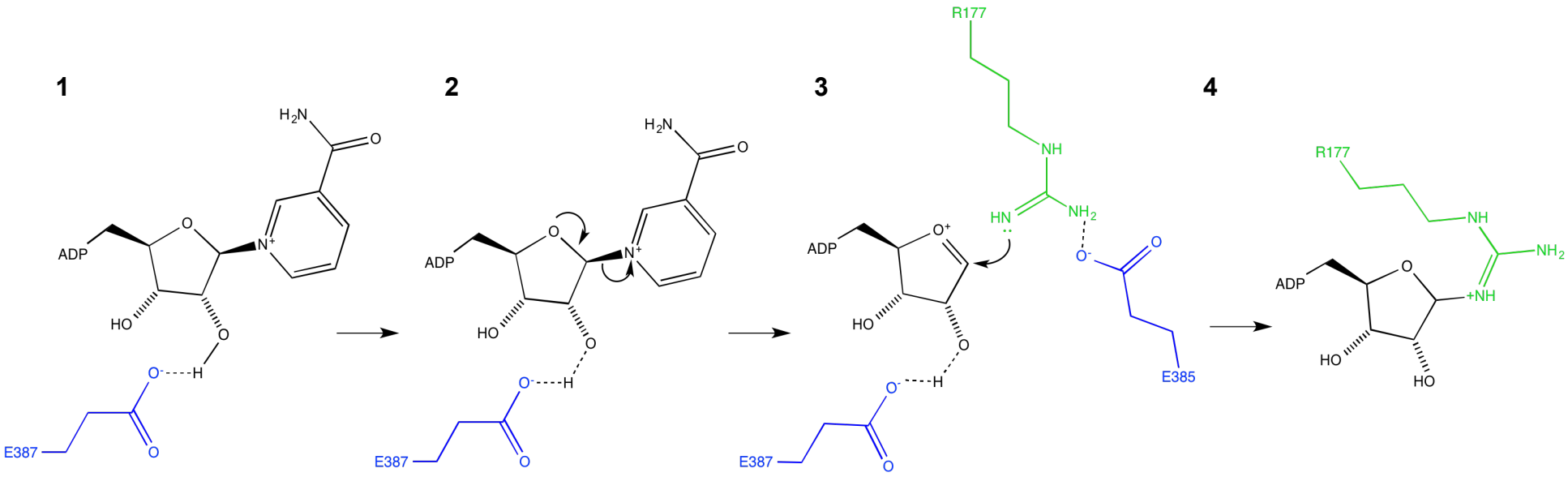


Figure 1.2.2c- Proposed mechanism of ADP-ribosylation of actin via an SN1 mechanism. 1: NAD (black) interacts with Glu-387 (blue) in the active site of CDTa. 2: A H-bond is formed between Glu-387 of CDTa (blue) and the 2'-OH group of the ribose ring (black). 3: An oxocarbenium intermediate is formed rendering the ribose group susceptible to nucleophilic attack and the ARTT-loop is rearranged so that Glu-385 of CDTa (blue) brings Arg-177 (green) of actin into the active site. 4: ADP-ribose is transferred onto Arg-177 (green) of actin. Image was created using ChemDraw ®.

1.3 Summary and PhD Objectives

Introductory Discussion with Primary Objectives

As discussed in the previous subchapters, the individual domains of the LCTs are well characterised however the full-length toxins are not well understood at a structural level. The LCTs tend to be produced by the more virulent strains of *C. difficile* in addition to the binary toxin, however there are strains of *C. difficile* that have variations in their ability to produce these toxins, and there is great uncertainty over the roles of each of these toxins individually in pathogenesis (Kuehne *et al.*, 2010; Lyerly *et al.*, 1985; Lyras *et al.*, 2009). As some of the most pathogenic strains of *C. difficile* produce CDT in addition to the LCTs, it is tempting to question whether or not CDT plays an adjunctive role to the LCTs in pathogenesis (Davies *et al.*, 2011).

In the previous section of this chapter, the structural aspects of the individual domains of the LCTs were discussed in detail. The full-length structures of the LCTs are yet to be determined, but the structures of some of the individual domains of these toxins have been solved. From the current structures there is now an improved understanding of the binding, the autoproteolysis and the mechanism by which the LCTs glucosylate Rho GTPases. These findings have opened up a number of avenues that are of great value for use in drug therapeutics. Determination of the crystal structures of the remaining unsolved domains is highly desirable, in addition to solving the structure of an enzyme–substrate complex. This will provide an understanding of how the toxins function as a whole. However, there is some ambiguity associated with the current domain structures and that uncertainty is that effectively these structures are small fragments of the full-length toxin. These individual domains could potentially behave differently when not associated with the remaining domains. Therefore, it is essential that the crystal structure of the full-length toxin is resolved. This feat would be particularly enlightening to discern the hydrophobic 'D' domain region, for which there are no high-resolution structures. One of the principal objectives for this PhD was to work alongside Public Health England towards elucidation of the crystal structure of the full-length Toxin A.

In terms of CDT, the objective was to test the hypothesis that the enzymatic component CDTa acts via an SN1 reaction, by performing site directed mutagenesis of active site residues. In addition, the intention was to characterize the transport component at a structural level, as currently there is only structural data for the enzymatic component. As it has been shown that CDTa and CDTb when combined are toxic to Vero cells, it would be beneficial to characterize the binary toxin as a whole complex in order to elucidate their interactions (Sundriyal *et al.*, 2010). This could provide a platform for highly specific drug design to block translocation of CDTa into target cells. Concurrently, it would also be of use to characterize the interactions between the CDTa component and its ADP-ribosylation target, actin. Obtaining a complex of CDTa-actin and solving the structure, would provide a target for drug design from a different perspective. By either blocking translocation of CDTa or blocking ADP-ribosylation of actin, highly specific drugs designed with the aid of complex structures could potentially inhibit the cytotoxic effect of CDT.

Chapter 2

Toxin A

2.1 Crystallisation of Toxin A

2.1.1 Introduction

Toxin A

Clostridium difficile produces two large exotoxins: Toxin A (308 kDa) and Toxin B (269.6 kDa), both of which are members of the Large Clostridial Toxins (LCT) family. These toxins have a multi-modular domain organisation, which is divided into four functional domains. These domains are represented by an ABCD model, see Figure 2.1.1a (Davies *et al.*, 2011; Jank and Aktories, 2008).

Figure 2.1.1a The ABCD domain organisation of Toxin A. The multi-modular arrangement of Toxin A is shown with the enzymatic glucosylating domain ('A' domain) in blue, the cysteine protease domain ('C' domain) in green, the hydrophobic pore forming domain ('D' domain) in magenta and the binding domain ('B' domain) in purple.

As discussed in detail in Chapter 1.2, Toxin A glucosylates Rho GTPases upon entry into target cells. Briefly, the 'B' domain interacts with carbohydrate receptors on the cell surface and the whole toxin is taken up into endosomes by receptor-mediated endocytosis. The toxin undergoes conformational changes upon the decreasing pH in the endosome. The 'D' domain forms a pore in the endosomal membrane allowing the translocation of the 'C' domain and 'A' domain into the cytosol. The 'C' domain then interacts with inositol hexakisphosphate ($InsP_6$), which leads to autocatalytic cleavage releasing the enzymatically active 'A' domain into the cytosol. The activated 'A' domain catalyses the transfer of cellular uridine diphosphate glucose (UDP-Glc) to the Thr-37 residue of Rho GTPases. This disrupts the interaction with effector molecules leading to disorder in the cytoskeleton and cell death.

The three-dimensional structures of full length Toxin A and Toxin B are yet to be elucidated. However, three of the four domains are structurally characterised (domains 'A', 'B' and 'C'), but the 'D' domain is still poorly understood. The aim was to crystallise the full length Toxin A and use X-ray crystallography to solve the structure.

Crystallisation

Protein crystals are highly ordered lattices made up of non-covalently bound protein molecules. As proteins tend to be irregular in shape, they do not make good 'building blocks' and so packing of proteins into lattices is a delicate process that is sensitive to many environmental factors, hence why crystallisation of proteins is often challenging.

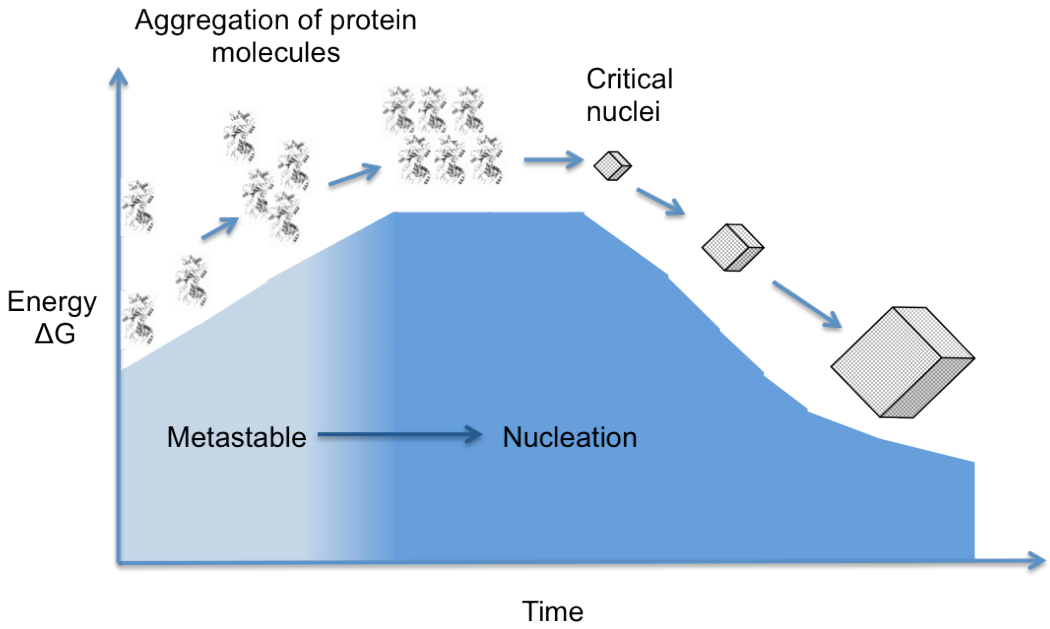


Figure 2.1.1b- Critical nuclei formation. Proteins must overcome an energy barrier to form crystals. Once the protein reaches the high-energy intermediate critical nuclei stage, crystals may grow (Bergfors, 2009).

When the correct conditions are met for protein crystallisation, the soluble protein undergoes controlled precipitation to form ordered crystals. The solubility of the protein is dependent on a number of factors, which include the

concentration of protein, the pH, the temperature and the concentration of any other additives such as salts or precipitants.

In order for a protein to crystallise it must overcome an energy barrier and enter a supersaturated state without precipitating completely, see Figure 2.1.1b. A schematic diagram to represent the relationship between concentration of protein and concentration of crystallising agent can be seen in Figure 2.1.1c (McPherson, 1989) . To obtain well ordered, diffraction quality crystals, the mechanism by which the protein enters a supersaturated state must be well controlled (Arakawa and Timasheff, 1985; Boistelle and Astier, 1988; Boistelle, 1986). A supersaturation state eventually gives rise to the formation of critical nuclei, which in some cases can seed crystal growth.

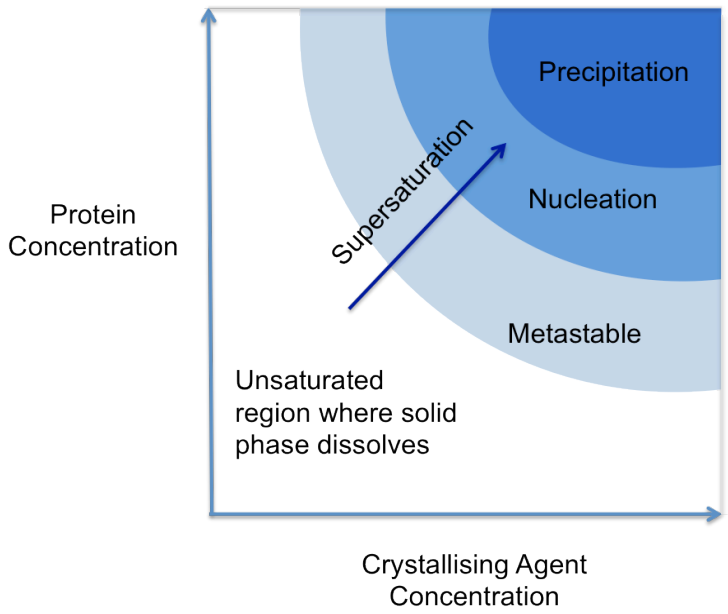


Figure 2.1.1c- Protein supersaturation curve. Crystallisation occurs in the nucleation zone under supersaturation conditions. However, protein can precipitate out of solution if this supersaturation is too high.

Traditionally techniques such as batch crystallisation and equilibrium dialysis were used for protein crystallisation, however the vapour diffusion method is now favoured. The protein and crystallising agent are mixed together, usually in a 1:1 ratio, in a small volume and are equilibrated against a large volume of the crystallising agent. Initially the droplet containing protein and

crystallising agent will have an insufficient concentration of precipitant to stimulate nucleation, but as vapour diffuses from the droplet to the reservoir, the concentration of precipitant gradually increases, leading to controlled movement from the metastable phase to nucleation. If this process happens too rapidly, precipitation of protein can occur, or alternatively if the crystallising agent conditions are not appropriate for that protein specifically, crystals will not grow (Blundell and Johnson, 1976).

There are two commonly used vapour diffusion techniques; hanging drop, where the droplet is suspended over the reservoir solution, and sitting drop, where the droplet sits in a well in close range to the reservoir solution, Figure 2.1.1d (Feigelson, 1988).

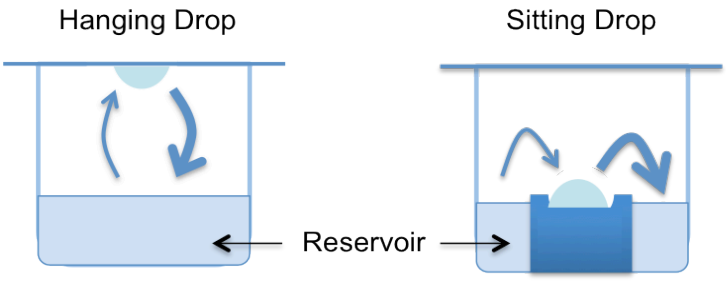


Figure 2.1.1d- Vapour diffusion methods. Hanging drop and sitting drop vapour diffusion methods are displayed, showing direction of diffusion.

There are two distinct phases of crystallisation, the first is the preliminary screening of conditions required for crystal growth, and the second is optimisation of these conditions to improve crystal size and quality. Once initial hits have been observed in the preliminary screening, these conditions can be optimised in attempts to improve the crystal quality. The standard technique of optimising conditions includes altering temperature, buffer pH, and varying salt or precipitant concentrations. In some cases, the crystals either cannot be reproduced or appear with the same morphology, showing no improvements. In such cases, various seeding techniques can be used, three of which will be discussed here. All seeding techniques share the same general principle; the seed provides the template and instigates the assembly of the protein molecules

to form crystals with the same characteristics as the crystal from which the seed originated.

Streak seeding: A probe, usually a cat's whisker, is used to touch an existing crystal to dislodge seeds from it. These seeds are then transferred into a pre-equilibrated droplet containing the same conditions the seed was taken from, by rapidly running the whisker in a line through the drop. This method should provide seed nucleated crystal growth along the line in the drop.

Microseeding: In this method, pre-existing crystals are crushed to create a seed stock. This stock is then diluted in series and re-introduced back into a pre-equilibrated droplet. This again should seed crystal growth.

Macroseeding: A suitable crystal of good morphology and free of defects is removed from a well and is washed repeatedly in stabilising solution. The crystal is then transferred to a pre-equilibrated drop where in this case the crystal is expected to grow in size rather than stimulate new crystal growth.

The aim of this subchapter (Chapter 2.1) was to use X-ray crystallography to solve the structure of the full-length Toxin A.

2.1.2 Materials and Methods

Purification

Pure Toxin A was kindly provided by Dr Clifford Shone, Public Health England (PHE). To ensure > 95% purity for crystallisation, a final cleaning step was performed. Toxin A aliquots were thawed on ice, centrifuged at 13,000 rpm to remove any precipitate and passed through a gel-filtration column; superdex-200 10/300 GL (GE Healthcare), pre-equilibrated with Toxin A buffer (50 mM Tris-HCl pH 7.5 and 150 mM NaCl). Collected fractions were analysed using 4-16 % Bis-Tris sodium-dodecyl-sulphate polyacrylamide-gel electrophoresis (SDS-PAGE) and those fractions containing pure Toxin A were pooled and concentrated to 10 mg/ml using 30 kDa cut-off concentrator tubes (Amicon ® Ultra, Millipore). The final purity of concentrated Toxin A was analysed by 4-16 % Bis-Tris SDS-PAGE.

Crystallisation

Preliminary Screening

A pre-crystallisation test (PCT) was carried out to determine the optimum protein concentration for crystallisation. Reagents displayed in Table 2.1.2a were made in house following the PCT guide (Hampton Research). A 4-well crystallisation plate was set up using the four reagents described in Table 2.1.2a. The PCT was set up using the hanging drop vapour diffusion method with a drop size of 1:1 µl ratio of protein to reservoir. The PCT plate was incubated at 16 °C for 30 min before being analysed using a light microscope. If the concentration of the protein was deemed suitable for crystallisation using the PCT guide, then the preliminary screens were performed. Preliminary screens were performed with the Phoenix Liquid Handling System (Art Robbins Instruments) using the sitting drop method with a ratio of 0.3:0.3 µl protein to reservoir ratio. Chemicals and reagents were sourced from Sigma-Aldrich unless stated otherwise.

Reagent 1	0.1 M Tris-HCl pH 8.5, 2.0 M Ammonium sulfate
Reagent 2	0.1 M Tris-HCl pH 8.5, 1.0 M Ammonium sulfate
Reagent 3	0.1 M Tris-HCl pH 8.5, 0.2 M Magnesium chloride hexahydrate and 30 % w/v Polyethylene glycol 4,000
Reagent 4	0.1 M Tris-HCl pH 8.5, 0.2 M Magnesium chloride hexahydrate and 15 % w/v Polyethylene glycol 4,000

Table 2.1.2a- Pre-crystallisation screen reagents used in test.

It was established that the concentration range 4-10 mg/ml was suitable for crystallisation using the PCT. Therefore, the following 96-well preliminary screens were performed with protein concentrations 4 mg/ml, 8 mg/ml and 10 mg/ml: Clear Strategy Screen I HT-96, Clear Strategy Screen II HT-96, JCSG-plus HT-96, Heavy + Light twin pack HT-96, Morpheus HT-96, MIDAS HT-96, MemGold HT-96, PGA screen HT-96, ProPlex HT-96, PACT premier HT-96, Stura Footprint Combination screen HT-96, and finally Structure screen I + II HT-96 (Molecular Dimensions). The preliminary screens performed were all incubated at 16 °C, and monitored initially after 24 hours followed by fortnightly thereafter.

Optimisation of Preliminary Hits

After observation of any crystal hits from the above preliminary screens, the corresponding crystallisation conditions were further tested to optimise crystal growth. Optimisations were performed on 24-well plates using either the hanging drop or sitting drop method, with a reservoir volume of either 0.5 ml or 1.0 ml. As discussed previously in Chapter 2.1.1, the crystallisation screens normally vary in buffer type and pH, salt type and concentration, and precipitant type and concentration. These various concoctions have been shown to stimulate or contribute to the growth of crystals. Temperature, protein concentration and crystallisation techniques are also known to affect crystal growth.

Initial crystallisation hits may often be too small, poorly diffracting or the crystals may dissolve easily, and thus quite often the initial crystallisation condition may require optimisation. Details on optimisation techniques have been discussed previously in Chapter 2.1.1, and further details on the specific crystallisation conditions will be discussed in the results section Chapter 2.1.3.

X-ray Diffraction Data Collection

Upon observation of Toxin A crystal hits, the conditions in which crystals were grown were examined to determine whether or not crystals should be treated with cryoprotectant. This will be discussed in the following subchapters where relevant. Crystals were then mounted in LithoLoops (Hampton Research), rapid frozen in liquid nitrogen and taken to Diamond Light Source to test for signs of diffraction. During preliminary testing of crystals the LithoLoops were mounted on the goniometer and three images were taken at rotation angles 45 ˚ apart with 1 second exposure. Further details on crystal treatments and data collection will be described in the results section, Chapter 2.1.3.

2.1.3 Results

Purification of Toxin A

Gel-Filtration of Toxin A

Nearly pure Toxin A was provided by PHE, however a final polishing step of gel-filtration was performed using a superdex-200 10/300 GL (GE Healthcare). Fractions of 500 µl were collected for the duration of the run and were analysed by 4-16 % Bis-Tris SDS-PAGE. As displayed in the chromatogram profile, Figure 2.1.3a, there is a sharp peak between approximately 8-12 ml (fraction A13-B11) and a small contaminant peak at 14.5 ml (fraction B6). When comparing this chromatogram profile with that of the calibration profile, it appears that Toxin A is eluted at a point at which is expected for proteins of approximately 600 kDa. This could suggest that Toxin A exists as a dimer in solution, or that the superdex-200 10/300 GL is unable to resolve high molecular weight proteins.

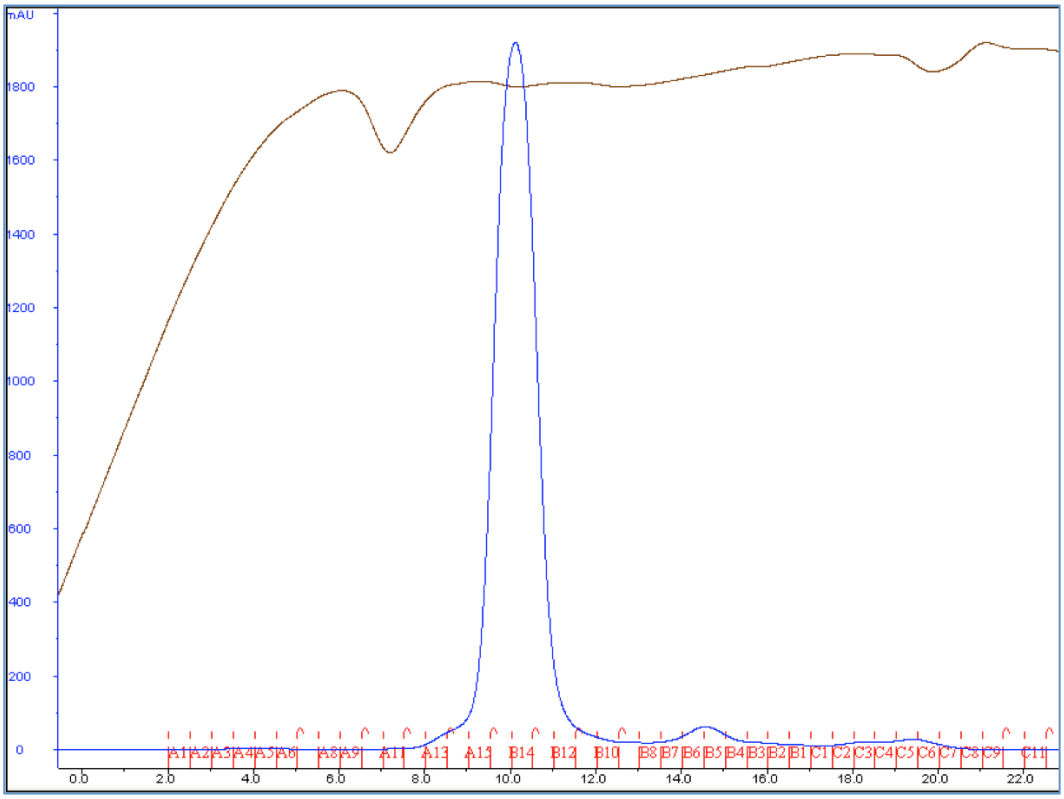


Figure 2.1.3a- An example of a Toxin A gel-filtration chromatogram using a superdex-200 10/300 GL column. Fractions A13-B11 contained pure Toxin A. The blue line represents absorbance at 280nm and the brown line represents conductivity.

SDS-PAGE Analysis of Purity

The purity of Toxin A was assessed using 4-16 % Bis-Tris SDS-PAGE, which was performed at 200 V for 1 hr. The fractions A13-B11 were pooled, concentrated and loaded onto the gel, where the final pure Toxin A can be seen in Figure 2.1.3.b. The SDS-PAGE results suggest that Toxin A is > 95 % pure.

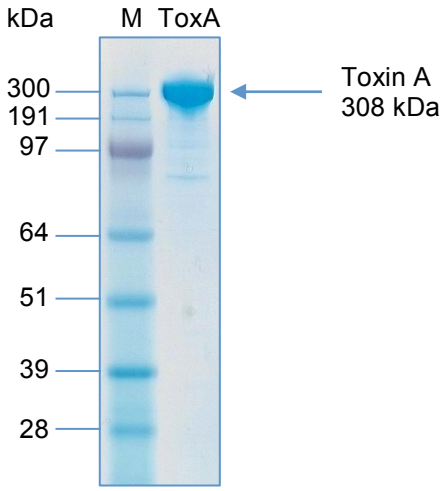


Figure 2.1.3.b- 4-16% Bis-Tris SDS-PAGE gel picture. The pre-stained protein ladder (SeeBlue® Invitrogen) is visible in lane 1, followed by the pure Toxin A in lane 2. Gel picture provided by Harriet Denton, Public Health England,

Crystallisation of Toxin A

Preliminary Hits

Preliminary screens for Toxin A were set up as discussed in Chapter 2.1.2, with a range of concentrations: 4 mg/ml, 8 mg/ml and 10 mg/ml. Tiny cube-like microcrystals were observed in two of the primary 96-well screens: two wells from the PGA HT-96 screen contained crystals in addition to two wells from the MemGold HT-96 screen, displayed in Figure 2.1.3.c. The crystals observed in the MemGold HT-96 and PGA HT-96 crystallisation screens were all small and cube-like. Their overall morphologies were similar but they differed quite

significantly in size, although they were all considered to be quite small at < 50 microns. However, it is clear from the pictures displayed in Figure 2.1.3c that the crystals in well D looked most promising.

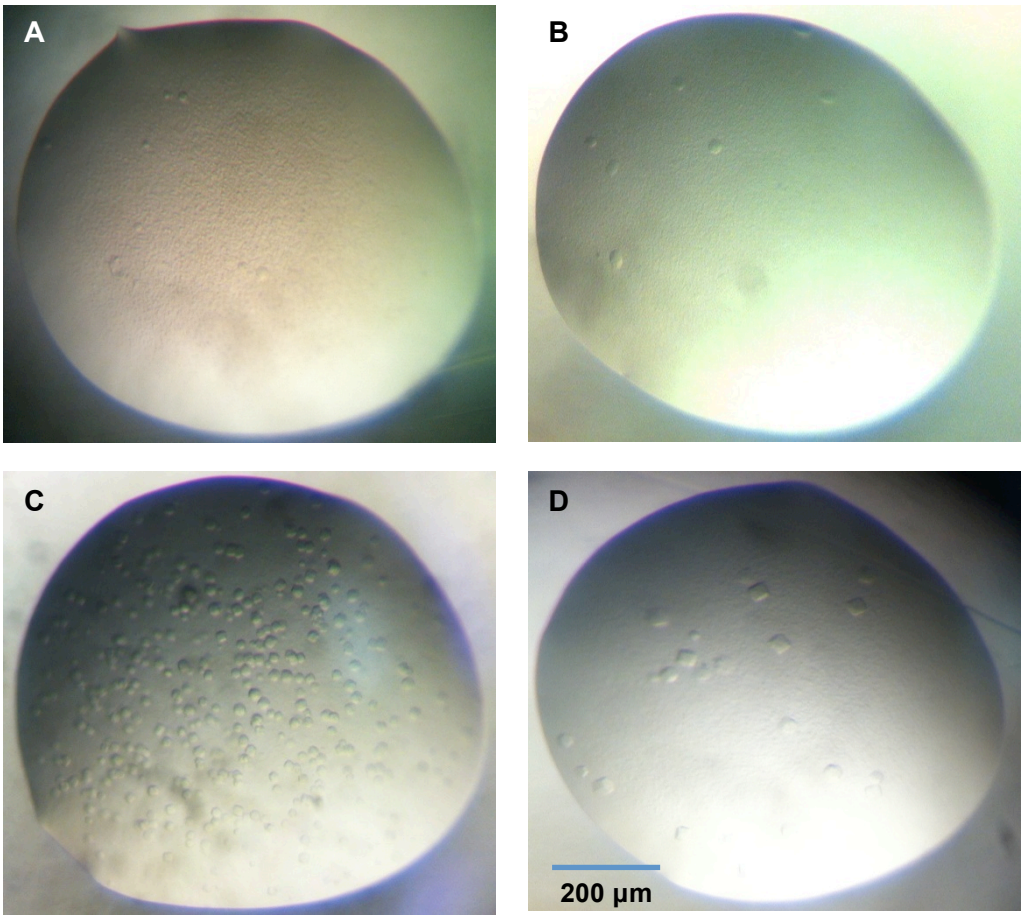


Figure 2.1.3c Preliminary crystallisation hits of Toxin A. Crystal hits were noted in the above four wells on the following 96-well plates: A and B: MemGold HT-96, C and D: PGA HT-96. Details of the crystallisation conditions are displayed in table 2.1.3a.

As discussed in Chapter 2.1.2 under '*Optimisation of preliminary hits*`, the crystallisation reservoir commonly contains a buffer, salt and precipitant. The crystallisation hits are shown in the above figure with the corresponding crystallisation conditions displayed in Table 2.1.3a. It is clear from the table that the most common features amongst the conditions are the presence of sodium and also low molecular weight polyethylene glycol (PEG) derivatives.

At least two crystals from each well from the initial hits were mounted in LithoLoops (Hampton Research) and taken to Diamond Light Source to test for signs of diffraction. No additional cryoprotectant was added due to the presence of PEG in the crystallisation conditions. For hits A and C from Figure 2.1.3c, mesh LithoLoops were used to ensure the crystals were scooped out of the wells successfully.

One at a time, the LithoLoops containing crystals were mounted on the goniometer and were rotated at 45 ˚ angles whilst being exposed to a beam of X-rays.

Crystal Hits From Figure 2.3.1.c	Protein Concentration	Screen and Well	Crystallisation Condition
A	8 mg/ml	MemGold HT-96 D2	0.05 M magnesium acetate 0.05 M sodium acetate pH 5.0 28% PEG 400
B	8 mg/ml	MemGold HT-96 G9	0.05 M sodium chloride 0.1 M sodium citrate pH 5.5 26% PEG 400
C	8 mg/ml	PGA HT-96 E10	0.1 M ammonium sulphate 0.3 M sodium formate 0.1 M sodium cacodylate pH 6.5 3% PGA-LM 20 % PEG 550 MME
D	8 mg/ml	PGA HT-96 H6	0.1 M ammonium sulphate 0.3 M sodium formate 0.1 M Tris pH 7.8 3% PGA-LM 20 % PEG 550 MME

Table 2.1.3a Crystallisation conditions for preliminary hits. Crystallisation conditions displayed for the corresponding crystal pictures displayed in Figure 2.1.3c.

No diffraction was observed for any of the mounted crystals from the wells pictured in Figure 2.1.3c. If the crystals had been salt we would have observed the characteristic large diffraction spots that are usually observed for salt crystals. The lack of diffraction spots indicates that the crystalline material could be

protein, rather than salt, but that the crystallisation condition needs further optimisation for diffraction studies.

Optimisation of Crystal Quality

Since the crystals from the preliminary hits showed no signs of diffraction, it was clear that the crystallisation conditions needed optimising. The attempts made at optimisation of crystal quality are displayed in Table 2.1.3b.

Screen	Stock Solution	Concentration Range Trialled
MemGold HT-96 D2	0.05 M magnesium acetate	0.05 M
	0.05 M sodium acetate pH 5.0	0.05 M
	28% PEG 400	0-30%
MemGold HT-96 G9	0.05 M sodium chloride	0.05 M
	0.1 M sodium citrate pH 5.5	0.1 M
	26% PEG 400	0-30%
PGA HT-96 E10	0.1 M ammonium sulphate	0.1 M
	0.3 M sodium formate	0.3 M
	0.1 M sodium cacodylate pH 6.5	0.1 M
	3% PGA-LM	0-8 %
	20 % PEG 550 MME	0-25 %
PGA HT-96 H6	0.1 M ammonium sulphate	0-0.3 M
	0.3 M sodium formate	0-0.5 M
	0.1 M Tris pH 7.8	0.1 M
	3% PGA-LM	0-8 %
	20 % PEG 550 MME	0-25 %

Table 2.1.3b Crystallisation conditions and optimisations. The crystal conditions are displayed in addition to the variations trialled in attempts to optimise the preliminary hits that are displayed in Figure 2.1.3c.

The stock solutions displayed in Table 2.1.3b were filtered using 0.4 µM pore size, 0.33 mm sterile filters (Millex) prior to use in crystallisation optimisation. The first round of optimisations were set up on a 24-well plate, at

concentration 4 mg/ml, with 1 ml reservoir volumes, 2 µl: 2 µl protein to reservoir drop size ratio, and were performed with the hanging drop method (see Chapter 2.1.1 for details on hanging drop methods). The optimisation plates were set up at room temperature but then incubated at 16 °C. The plates were checked after 24 hours and then weekly thereafter.

No crystals were observed on either of the two MemGold HT-96 optimisation plates, nor the PGA HT-96 E10 plate. A small number of cube like crystals were observed from the PGA HT-96 H6 optimisation plate, in the original crystallisation condition, however a number of crystals were also noted in wells containing a reduced sodium formate concentration of 0.2 M, see Figure 2.1.3d. The precipitant concentration should be kept constant with the original crystallisation condition as no crystals were observed where these concentrations were varied.

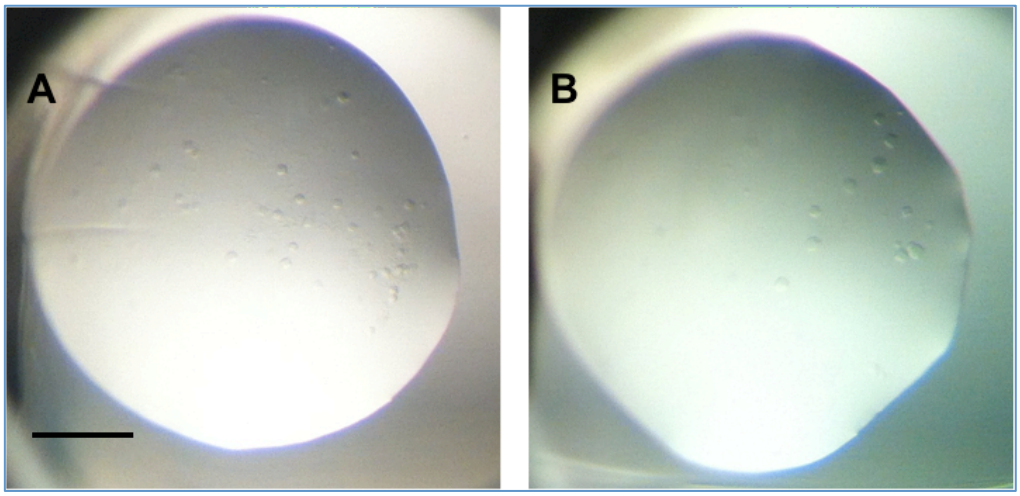


Figure 2.1.3d- Crystals of Toxin A from optimisation trials. A: Crystals grown in original PGA HT-96 H6 condition on optimisation plate. B: crystals grown in reduced sodium formate concentration. The scale bar in B represents 200 µm.

Using the same method as described for the initial crystallisation hits, crystals from each well of the optimisation hits were mounted in mesh LithoLoops (Hampton Research) and taken to Diamond Light Source to test for signs of diffraction. One at a time, the mesh LithoLoops containing crystals were mounted on the goniometer and were rotated whilst being exposed to a beam of X-rays. No diffraction was observed for any of the mounted crystals.

As the 'optimised' crystals showed no signs of diffraction, seeding techniques were employed using the crystals displayed in Figure 2.1.3d.

1. Initially macroseeding was attempted, whereby crystals from the PGA H6 optimisation plate were transferred into fresh drops on a new 24-well plate, containing the exact same reservoir condition and protein concentration that had been pre-equilibrated for 5 hours prior to seeding. Plates were incubated at 16 °C and checked after 24 hours and then weekly thereafter. After two months the macroseeded crystals showed no signs of improvement in terms of size or diffraction.

2. Microseeding was also attempted, whereby the 4 µl drop containing the protein crystals was resuspended in 10 µl of reservoir. The crystals were crushed using vortexing and diluted 10-fold, 100-fold, 1000-fold and 10,000-fold. A fresh 24-well plate was set up using the exact PGA H6 crystallisation condition including, drop size, protein concentration and reservoir volume. The dilution series of blended crystals (seeds) were then reintroduced into the pre-equilibrated 24-well plate; 1 µl of each of the diluted samples were added to the 4 µl drops on the 24-well plate. Plates were then incubated at 16 °C and checked after 24 hours and then weekly thereafter. No signs of crystal growth were observed, even after 3 months.

3. Finally, streak seeding was employed, whereby a cat's whisker pre-soaked in the PGA H6 reservoir was dipped into the drop containing crystals and was moved around in order to make contact with the crystals. As with the other seeding techniques, a fresh 24-well plate of PGA H6 crystallisation conditions was prepared in advance with the corresponding protein concentration and drop size. The cat's whisker, carrying crystal seeds, was then immersed into a fresh drop in a straight line (streak). Plates were incubated at 16 °C and checked after 24 hours. Crystals were observed after 24 hours, forming a line across the centre of the drop, precisely where the whisker had passed through, see Figure 2.1.3e.

The crystals grown from the streak-seeded plate showed a slightly different morphology from those previously observed, with a more elongated shape. Crystals shown in Figure 2.1.3e were mounted in mesh LithoLoops and were exposed to a beam of X-rays at Diamond Light Source. No diffraction was detected for any of the mounted crystals.

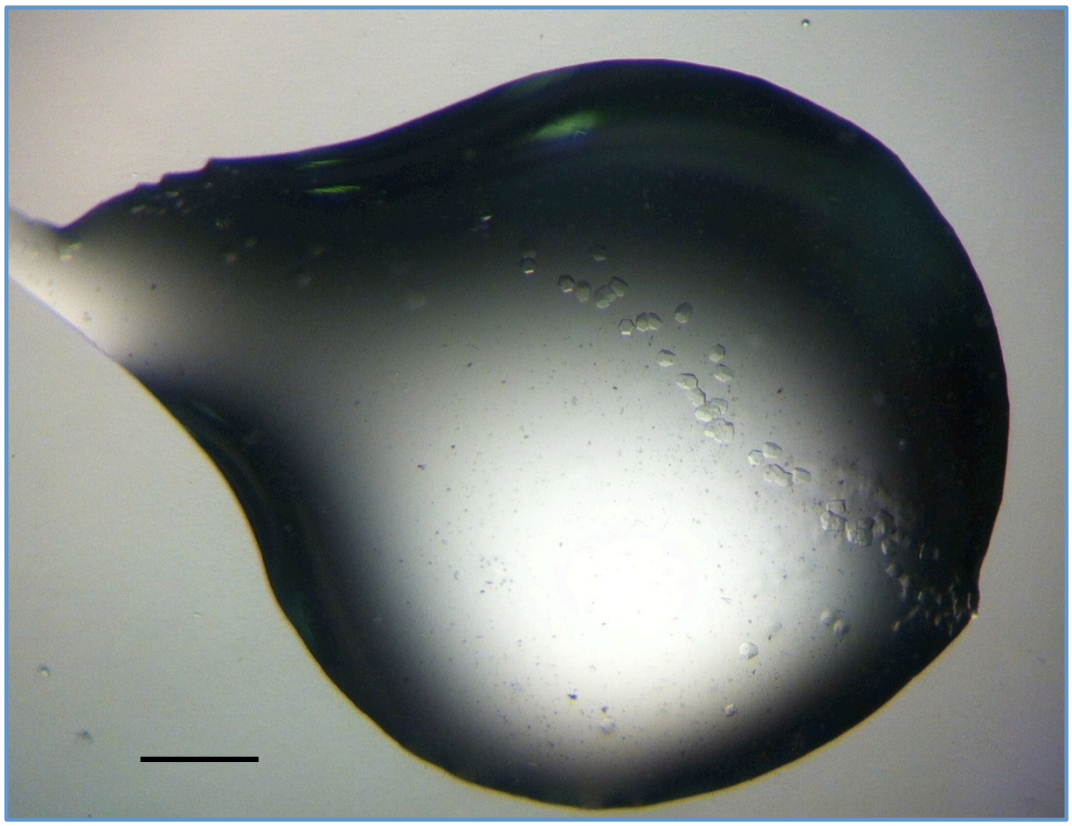


Figure 2.1.3e- Streak seed for Toxin A. Crystals of a slightly different morphology grew along the line that was seeded using a cat's whisker. The scale bar represents 200 µm.

As there had been no success up until this point at improving crystal quality, the preliminary screens were repeated using a slightly different approach. Toxin A was exchanged into different buffers and the preliminary screens were repeated. It was decided that if the buffer type was varied along with pH, this would extend the number of different conditions the protein would be exposed to.

Buffer Exchange of Toxin A

In order to maximise the chances of crystal growth in the preliminary screens, Toxin A, supplied by PHE was buffer exchanged into various other pH buffers. Buffer exchanges were performed on a superdex-200 10/300 GL column (GE Healthcare). The following buffers were chosen based on their pH range:

1. 50 mM MES pH 6.0 and 150 mM NaCl (Figure 2.1.3f-A)
2. 50 mM HEPES pH 7.0 and 150 mM NaCl (Figure 2.1.3f-B)
3. 50 mM Bicine pH 8.0 and 150 mM NaCl (Figure 2.1.3f-C)

The buffer exchange chromatograms displayed in Figure 2.1.3f show similar size exclusion profiles to the original pH 7.5 buffer profile. The protein elutes from the gel-filtration column at fraction A13, which is at 9 ml, and is consistent throughout all of the different pH buffers. As discussed earlier, according to the standard proteins calibration profile, Toxin A is eluting at a size twice of that to be expected. This is apparent with the entire range of buffers used; again suggesting Toxin A exists as a dimer.

Analysis of Buffer Exchanges using SDS-PAGE

Fraction B14 from each of the buffer exchange chromatograms were treated with SDS-PAGE reduced loading dye and were analysed by 8 % Tris-Glycine SDS-PAGE, which was performed at 200 V for 1 hour, see Figure 2.1.3g.

As shown in the SDS-PAGE stained gel picture, Toxin A appears in both monomeric and oligomeric forms. Although there are limitations when using 8% SDS-PAGE to analyse the size of the toxin, as it is a large macromolecule, it is clear that there is monomeric toxin visible on the gel and an upper band, of greater molecular mass. The gel-filtration chromatogram profile indicates that the protein is eluting at a point where the control protein of 600 kDa elutes, therefore as Toxin A is 308 kDa we can predict that this upper band is Toxin A in dimeric form.

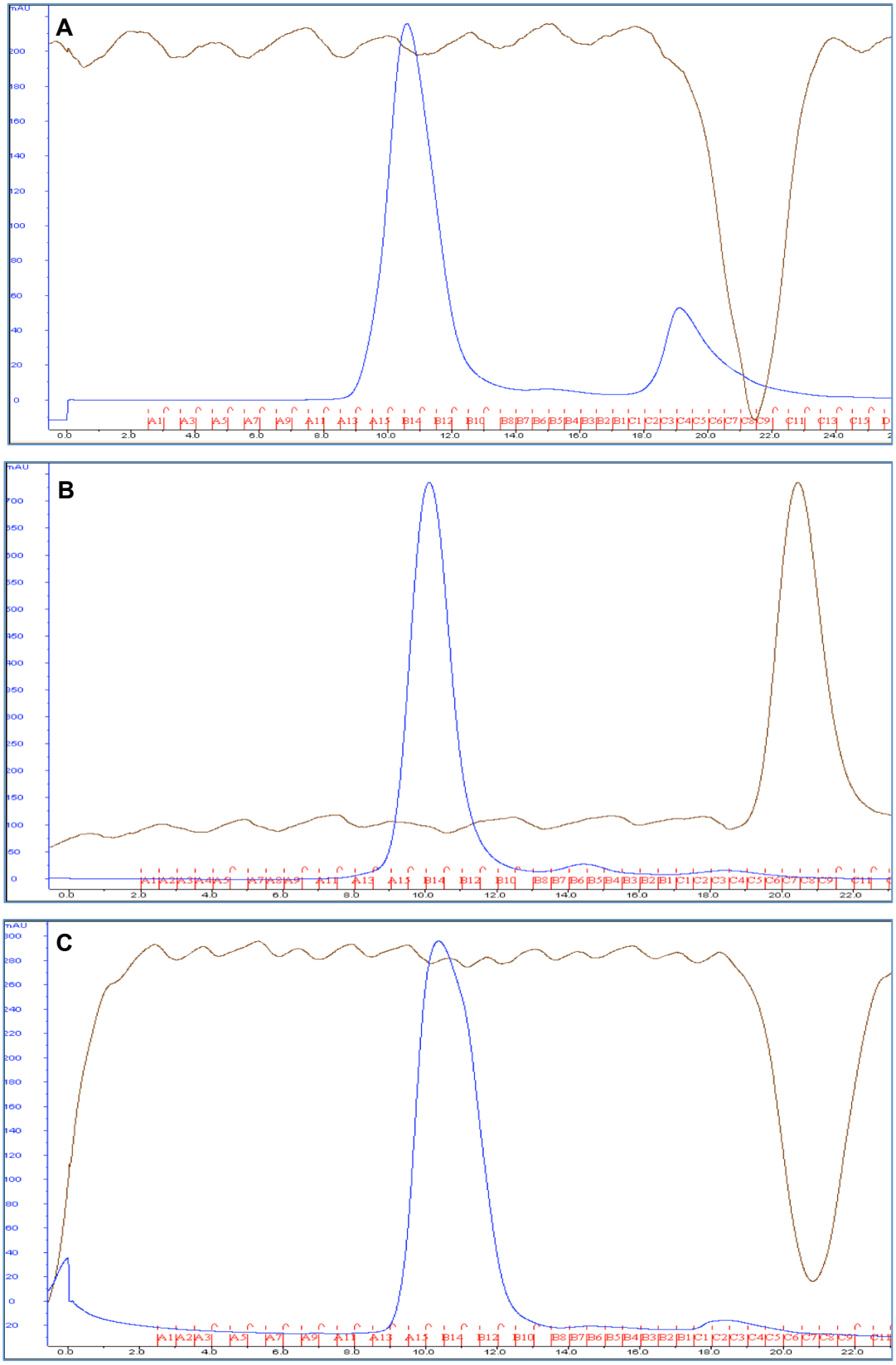


Figure 2.1.3f- Gel-filtration chromatogram profiles for buffer exchanges. **For all chromatograms, the blue line represents absorbance at 280nm and the brown line represents conductivity. A: Buffer exchange into 50 mM MES pH 6.0 and 150 mM NaCl. B: Buffer exchange into 50 mM HEPES pH 7.0 and 150 mM NaCl. C: Buffer exchange into 50 mM Bicine pH 8.0 and 150 mM NaCl**

The following 96-well preliminary screens were performed for Toxin A in the three different buffers at protein concentrations 4 mg/ml, on intelli 3-well plates (Molecular Dimensions): Clear Strategy Screen I HT-96, Clear Strategy Screen II HT-96, JCSG-plus HT-96, Heavy + Light twin pack HT-96, Morpheus HT-96, MIDAS HT-96, MemGold HT-96, PGA screen HT-96, ProPlex HT-96, PACT premier HT-96, Stura Footprint Combination screen HT-96, and finally Structure screen I + II HT-96 (All screens purchased from Molecular Dimensions). The crystallisation screens were all incubated at 16 °C, and monitored initially after 24 hours and then fortnightly thereafter.

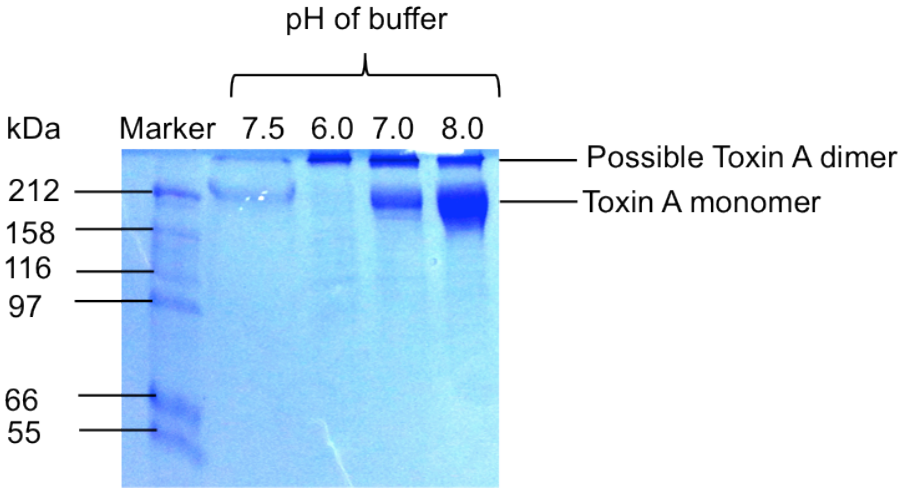


Figure 2.1.3g- 8% polyacrylamide Tris-Glycine SDS-PAGE displaying Toxin A in four different buffers. At buffer pH 6.0, it appears that Toxin A may be in it's dimeric form whereas in pH 7.0 and more so in pH 8.0 Toxin A is in both monomeric and dimeric forms.

Crystallisation hits were observed in the PGA HT-96 screen, well position H6 for pH 6.0 and pH 7.0, which as discussed earlier, was a hit for the pH 7.5 preliminary screens, see Figure 2.1.3h. The crystals have a beautiful elongated hexagonal-like morphology. There were no other hits observed for any other screens. It is my opinion that Toxin A is reluctant to crystallise perhaps due to the hydrophobic nature of approximately 50 % of the molecule. It appears that PGA HT-96 H6 is certainly the only condition in which this protein crystallises. These crystals were again tested for diffraction at Diamond Light Source, but as with previous crystals, no diffraction was observed.

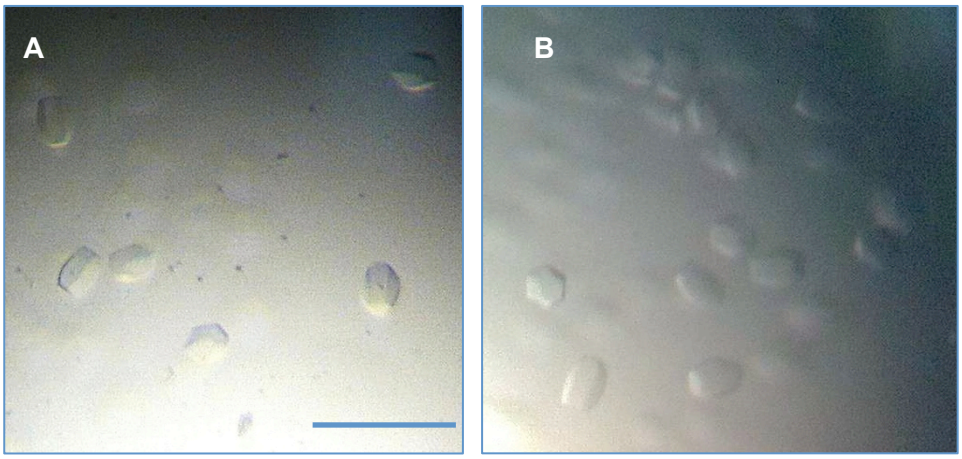


Figure 2.1.3h- Toxin A crystals grown in PGA HT-96 H6 crystallisation condition. A: Toxin A crystals from buffer 50 mM MES pH 6.0 and 150 mM NaCl. B: Toxin A crystals from buffer 50 mM HEPES pH 7.0 and 150 mM NaCl. The scale bar represents 100 µm.

After buffer exchanging Toxin A into the alternative pH buffers, and repeating the preliminary crystallisation screens, further attempts were directly focused around the PGA HT-96 H6 condition. In previous optimisation attempts, extensive phase separation was noted in all of the drops that were set up by hand. Therefore, 96-well plates were set up with the PGA HT-96 H6 crystallisation solution, using a Phoenix Liquid Handling System. The trials were set up on intelli 3-well crystallisation plates for each of the different pH solutions, at 4 mg/ml, 8 mg/ml and 10 mg/ml, using the sitting drop method with a ratio of 0.3:0.3 µl protein to reservoir ratio. The plates were then incubated at 16 °C, and monitored for crystal growth after 24 hours and weekly thereafter.

Diffraction of Toxin A Crystals

Early Signs of Diffraction

After 3 months, there were three large crystals observed in separate wells on the intelli 3-well plate, see Figure 2.1.3i. Protein in buffer 50 mM Tris-HCl pH 7.5 and 150 mM NaCl, formed crystals in PGA HT-96 H6 at concentration 8 mg/ml, and showed improved crystal morphology from previously grown crystals. These crystals have more defined edges and are elongated hexagonal in shape.

A cross-sectional view through the crystal reveals a perfect hexagonal shape, see Figure 2.1.3i-C. The unusual aspect of this protein is that the crystals do not grow consistently. Out of ninety-six wells of identical crystallisation conditions, only three wells on the plate contained crystals.

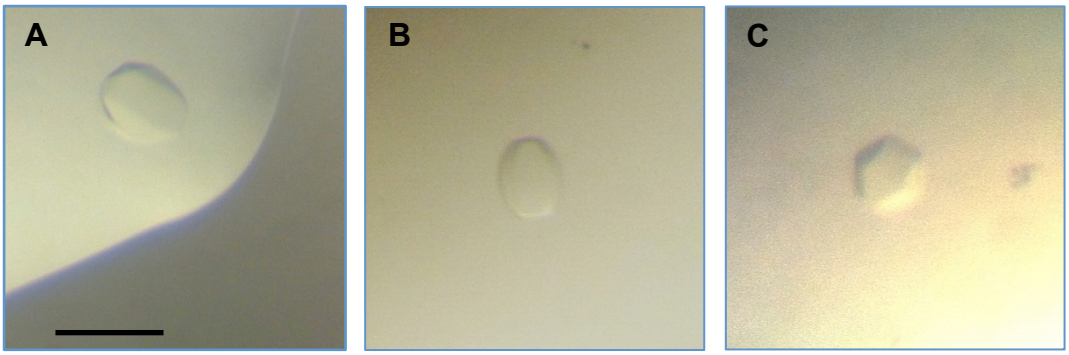


Figure 2.1.3i- Crystals of Toxin A in buffer 50 mM Tris-HCl pH 7.5 and 150 mM NaCl, grown in crystallisation condition PGA HT-96 H6. Crystals show improved crystal morphology to previously grown crystals at lower protein concentration. The scale bar represents 100 µm.

Crystals A and B shown in Figure 2.1.3i were individually mounted into LithoLoops (Hampton Research) and taken to Diamond Light Source to test for signs of diffraction. As crystals have shown no diffraction in the past, it was assumed that the protein molecules are not tightly packed, leading to the lack of diffraction. Therefore crystal C from Figure 2.1.3i was treated with a cross-linking agent, glutaraldehyde, which in theory should stabilise any interactions within the protein crystal. The reservoir for the chosen crystal was poisoned with 2 % glutaraldehyde and was resealed for 5 min. The crystal was then mounted into a LithoLoop (Hampton Research) and taken to Diamond Light Source to test the effect of glutaraldehyde on diffraction. One at a time, the LithoLoops containing crystals were mounted on a goniometer and were rotated at 45 ° angles whilst being exposed to a beam of X-rays. Crystals A and B from Figure 2.1.3i showed no signs of diffraction, but the hexagonal crystal, treated with glutaraldehyde showed the first diffraction spots ever detected for Toxin A. The crystal only diffracted up to a resolution of 25 Å, see Figure 2.1.3j, but it is a major step forward towards structure determination for Toxin A using X-ray crystallography.

Unfortunately, shortly after exposure to the beam, the crystal died and no longer diffracted, therefore a dataset could not be collected.

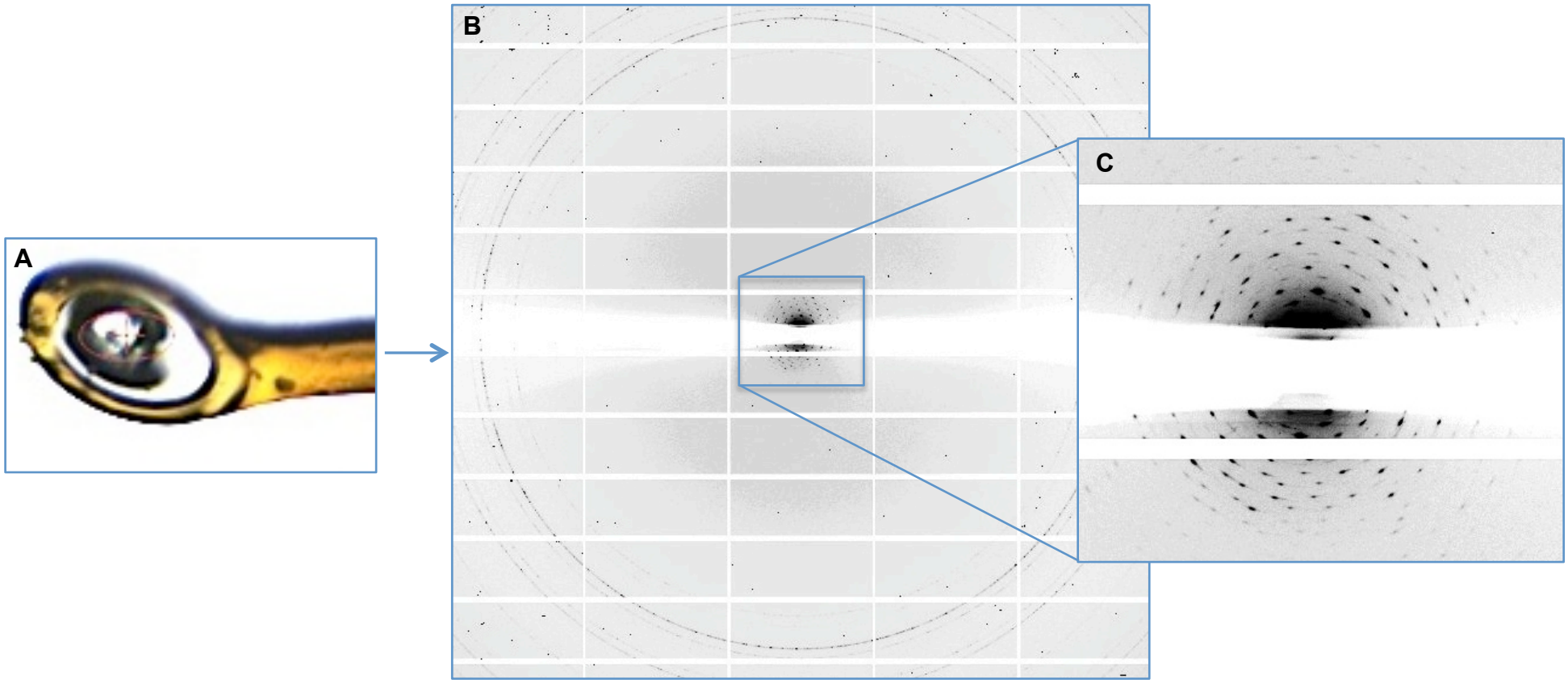


Figure 2.1.3j- First diffraction spots observed for Toxin A. A: Image of glutaraldehyde treated crystal in LithLoop during exposure to beam. Red circle and cross indicate beam position. B: Diffraction image of Toxin A. C: Zoomed in image of Toxin A diffraction spots.

Improved Diffraction and Dataset for Toxin A

After the first diffraction spots were observed for the PGA H6 crystals treated with glutaraldehyde, multiple plates were set up using the PGA H6 crystallisation condition (purchased from Molecular Dimensions) on both 96-well plates and 24-well plates. The plates were set up at 8 mg/ml and 10 mg/ml, at 4 ˚C, 16 ˚C and 21 ˚C. In addition, a 3:1 molar ratio of UDP-glucose to protein was used and a 3:1 molar ratio of manganese chloride to protein. This is because the glucosyltransferase 'A' domain of Toxin A binds UDP-glucose in the presence of divalent Mn^{2+} cations. Previous studies have shown the 'A' domain of Toxin A will only crystallise in the presence of both UDP-glucose and manganese (D'Urzo *et al.*, 2012). Therefore, both were added to the protein prior to crystallisation trials to determine whether or not they affect crystal growth and/or crystal quality. Within 1 week, crystals were observed in one well on a 96-well plate at 21˚C, with protein at 10 mg/ml containing UDP-glucose and manganese, see Figure 2.1.3k, which is the fastest crystal growth noted.

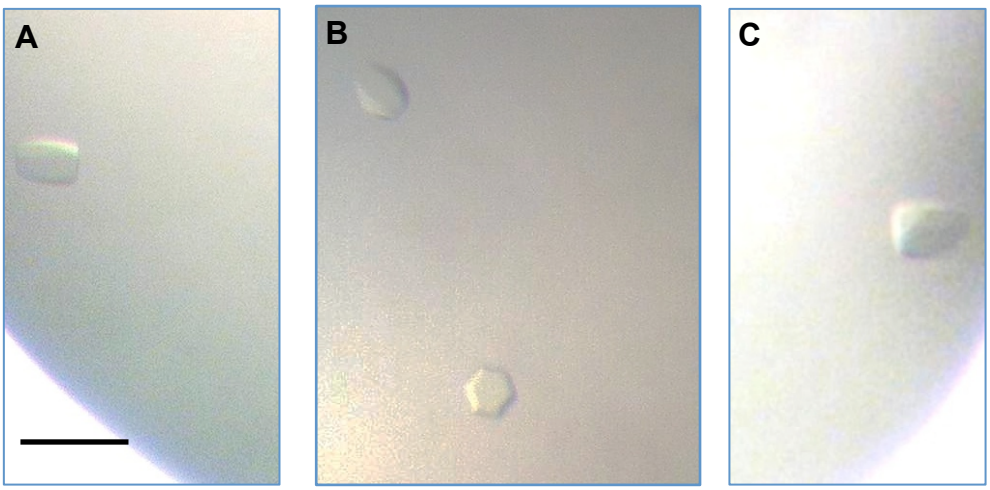


Figure 2.1.3k- Crystals of Toxin A grown at 10 mg/ml, 21˚C, with protein buffer containing UDP-glucose and manganese chloride. All crystals shown are from the same drop. The scale bar represents 100 µm.

The new crystals share a similar morphology to the previous diffracting crystals, with elongated shapes, and when viewed down one angle look perfectly

hexagonal. In order to improve diffraction, a number of methods were employed: Firstly crystal A from Figure 2.1.3k was washed and transferred into a new well containing a 10 % increase in PEG concentration, in the hope this would provide some cryo-protection and crystal stabilization, however the crystal dissolved. Secondly, the hexagonal shaped crystal in picture B was mounted directly in a LithoLoop with no alterations performed. Finally, the remaining crystal in picture B and that in picture C were treated with 2 % glutaraldehyde (in the reservoir). Crystal B was mounted in a loop and rapidly frozen after 5 min, and crystal C was mounted and frozen after 30 min.

Data Collection Statistics	**No glutaraldehyde (dataset 1)**	**5 min of glutaraldehyde treatment (dataset 2)**	**30 min of glutaraldehyde treatment (dataset 3)**
Space Group	$P3_121$	C2	$P3_121$
Cell Dimensions	a=400Å b=400 Å c=430Å α=90° β=90° γ=120°	a=714Å b=425Å c=443Å α=90° β=90.05° γ=90°	a=400Å b=400 Å c=433Å α=90° β=90° γ=120°
Resolution Range	69.05-10.72 Å	207.71-13.23 Å	79.97-11.34 Å
R_{symm}	0.20	0.22	0.18
I/σI (outer shell)	4.9 (2.1)	2.8 (1.9)	4.5 (2.3)
Completeness (outer shell)	82.1 (84.3)%	51.9 (53.8)%	90.5 (93.0)%
Total Number of Reflections	45812	28440	39750
Redundancy	3.2	1.8	2.9

R_{symm} = $\Sigma h \Sigma i[|Ii(h) - \langle I(h)\rangle|/\Sigma h \Sigma i Ii(h)]$ where *Ii* is the *ith* measurement, and $\langle I(h)\rangle$ is the weighted mean of all measurements of *I(h)*.

Table 2.1.3c- Data collection statistics for Toxin A. Displayed in the table are the data statistics for the three data sets collected for crystals displayed in Figure 2.1.3k. Data was processed using XDS and table statistics were generated by POINTLESS (Evans, 2006; Kabsch, 2010)**.**

The crystals were taken to Diamond Light Source and the usual test images were obtained at 45 ˚ intervals at the i04 beamline. All three crystals showed improved diffraction between 10.5-11 Å resolution, see Figure 2.1.3k. For each crystal data were collected at wavelength (λ) 0.9795 Å, 80 % transmission, with distance from the detector set to 1 m. At each oscillation of 1˚, the crystals were exposed for 10 s and for each crystal 50 images were collected. The data were auto processed and indexed using XDS to provide cell dimensions and were analysed using POINTLESS (Evans, 2006; Kabsch, 2010). The data were then scaled using SCALA (Evans, 2006). Table 2.1.3c displays the data collection statistics for each of the datasets collected from three different crystals.

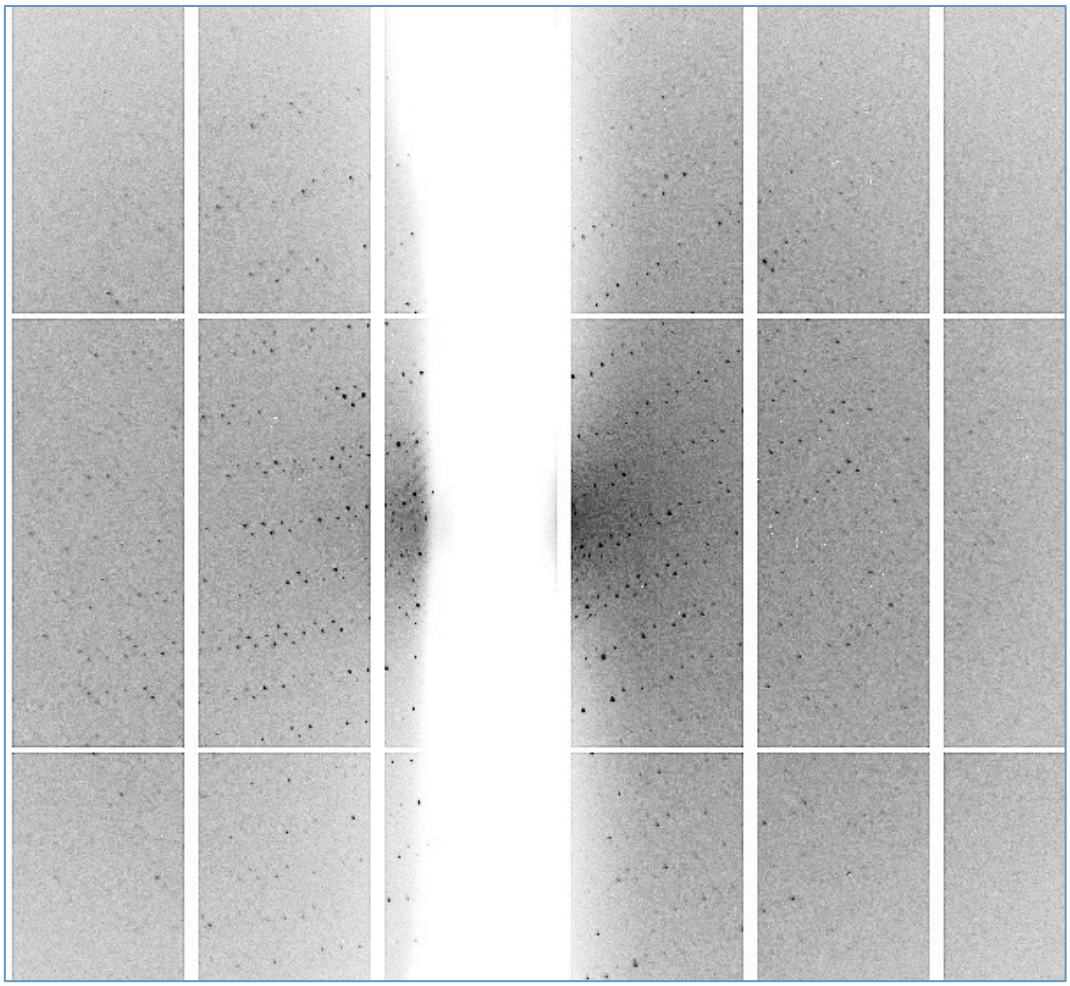


Figure 2.1.3l- Improved Toxin A diffraction pattern displaying well ordered lunes, with spots observed up to 10 Å.

Although an ambitious task, attempts were made at determination of the three-dimensional structure of Toxin A using both datasets 1 and 3. Intensive rounds of molecular replacement were performed using both Phaser and Molrep from the CCP4 Program Suite (McCoy *et al.*, 2007; Vagin and Teplyakov, 1997). The various fragments of Toxin A and Toxin B were used as search models for molecular replacement including: 2BVL, 4DMW, 3HO6, 3PEE, 2G7C and 2QJ6 (PDB codes) (D'Urzo *et al.*, 2012; Pruitt *et al.*, 2009; Reinert *et al.*, 2005; Shen *et al.*, 2011). Initially, the structures of the 'A' domain, 'B' domain and 'C' domain of Toxin A (PDB 4DMW, 2QJ6 and 3HO6 respectively) were input at the same time to provide the optimal amount of scattering data possible for molecular replacement. While the programs provided solutions, for a potential 'A' domain, the statistics were poor and there were multiple clashes between molecules. As expected the electron density map was of poor quality due to low-resolution data. After rigid body and restrained body refinement using Refmac (Murshudov *et al.*, 1997), the density showed no improvement. As an alternative method, multiple combinations of all of the known LCT domain structures were input into the molecular replacement software, and various other input factors were varied including 'number of molecules to search for'. The input models were also all used individually, searching for 1 molecule and 12 molecules separately, but again with no success.

There are various analytical statistics that can be performed even though a structure solution was not achieved. The Matthews Coefficient and solvent content were calculated from the unit cell and the molecular weight of the molecules in the unit cell, for each dataset.

Dataset	Number of molecules	Probability	Vm ($Å^3$/Da)	Solvent (%)
1	12	0.08	2.7	54.2
2	48	0.07	2.3	45.9
3	12	0.15	2.7	54.5

Table 2.1.3d- Matthew's coefficient and solvent content for the 3 datasets.

Given the size of the cell dimensions its not surprising that a large number of molecules in the asymmetric unit is predicted. However, as there were two datasets collected that were consistent with each other in terms of cell dimensions and space group (dataset 1 and 3), we can expect there to be 12 molecules in the asymmetric unit, or potentially 6 Toxin A dimers.

It may be ambitious to solve the structure of Toxin A at this level of resolution, especially due to its size and the lack of a model input for up to half of the protein. The crystallographic structures used as search models make up very small portions of the overall toxin. Therefore, higher to medium resolution data will be required in order to solve the structure by molecular replacement.

2.2 Low-Resolution Structure of Toxin A

2.2.1 Introduction

Toxin A

Toxin A is a member of the Large Clostridial Toxin family of proteins, as it is produced by *C. difficile* and has a mass of 308 kDa. The ABCD model accurately describes the domain organisation of Toxin A, see Figure 2.1.1a for details of the arrangement: the enzymatic domain (A), the binding domain (B), the cysteine protease domain (C) and the hydrophobic domain (D) (Jank and Aktories, 2008). Toxin A glucosylates Rho-GTPases which leads to cell rounding and cell death, as discussed in detail in Chapter 1.2. At present there is no high-resolution structure of the full length Toxin A, potentially because it is a difficult protein to work with, due to its size and nature. However, much effort has been spent solving the structures of the individual domains.

Electron Microscopy Data for Toxin A

Due to the lack of crystallographic data for Toxin A and Toxin B, some research groups have been working on alternative techniques to gain insight into the structural characteristics of these toxins. One such technique is electron microscopy. Pruitt *et al.* have imaged both Toxin A and Toxin B using negative stain electron microscopy, resulting in identification of the overall shape of the toxins, and have extended their analysis by modelling and mapping the domains of Toxin A with the electron microscopy images (Pruitt *et al.*, 2010). The electron microscopy images of Toxin A reveal homogeneous particles of non-symmetrical shape; see Figure 2.2.1a-A, whereas with Toxin B there was more of a heterogeneous selection of particles (data not shown).

Both Toxin A and B share structural similarities with a well-resolved globular head domain, a short inner tail domain and an extended long tail domain

(two tails in the case of Toxin B). After averaging a large number of particles, to give an overall particle shape, see Figure 2.2.1a-B, a 3D reconstruction was performed using the random conical tilt approach with the final result displayed as a mesh surface in Figure 2.2.1a-C. After imaging the individual domains by electron microscopy, monitoring pH effects on conformation, and using antibody labelling, it was concluded that the globular head domain represents the hydrophobic 'D' domain, the long extended tail represents the binding 'B' domain and the shorter tail represents the glucosyltransferase 'A' domain. The cysteine protease 'C' domain could not be located. A model of the binding 'B' domain was placed into the electron microscopy 3D map along with the crystal structures of the 'A' domain of Toxin B and the 'C' domain of Toxin A, to give an indication as to where these domains are organised in the full-length structure, see Figure 2.2.1a-C (Pruitt *et al.*, 2010).

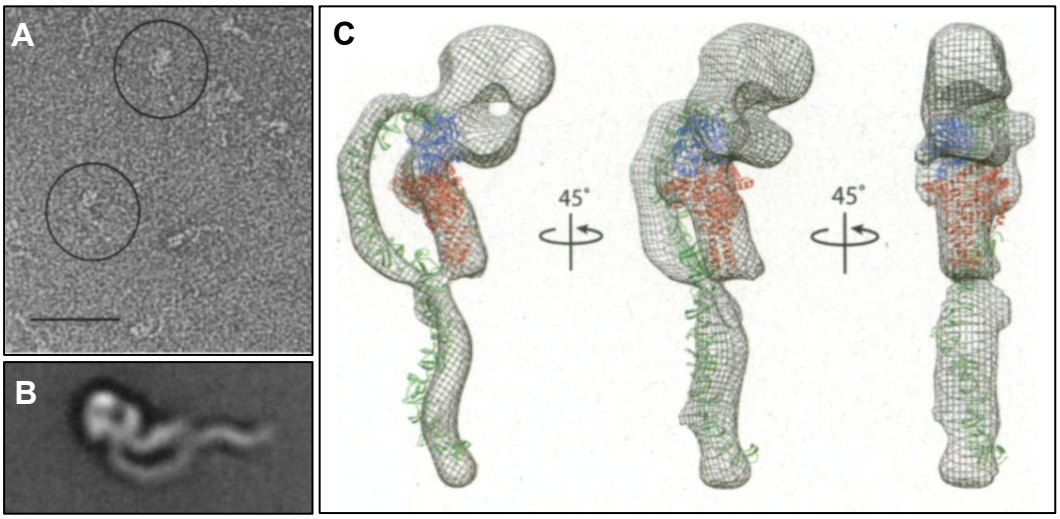


Figure 2.2.1a- Electron microscopy images of Toxin A and the molecular modelling of domains in a predicted 3D model of the full length toxin. A: Negative stain electron microscopy image of Toxin A molecules, with two molecules of Toxin A circled. The scale bar represents 500 Å. B: A representation of Toxin A created by averaging approximately 5000 Toxin A molecules from the electron microscopy images taken. C: A 3D reconstruction of the Toxin A molecule filtered to 25 Å displayed as a mesh surface, with known structural domains docked into the density. These known domains include the 'A' domain of Toxin B, the 'C' domain from Toxin A and a model of the 'B' domain from Toxin A. All images adapted and modified from Pruitt *et al.* 2010 (Pruitt *et al.*, 2010).

Small-Angle X-ray Scattering

Small-angle X-ray scattering (SAXS) is a technique used to determine low-resolution 3D structures of macromolecules in solution, see Figure 2.2.1b, which can provide information about the size, distribution and overall shape of the macromolecule.

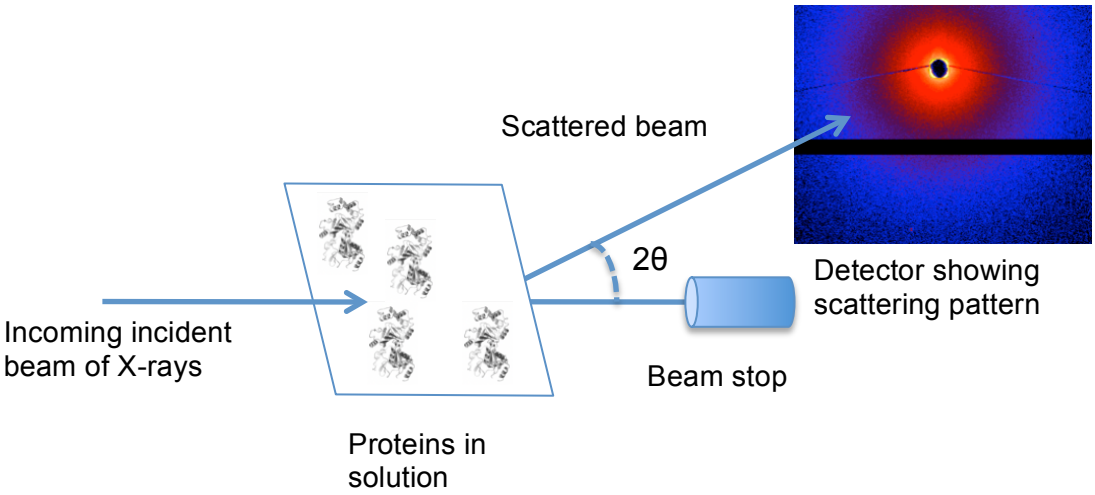


Figure 2.2.1b- Small-angle X-ray scattering. Adapted from Jacques and Trewhella, 2010 (

Figure 2.2.1c displays a flow diagram illustrating the steps taken towards obtaining a SAXS structure starting from expression and purification of the protein right through to the generation of an *ab initio* SAXS model.

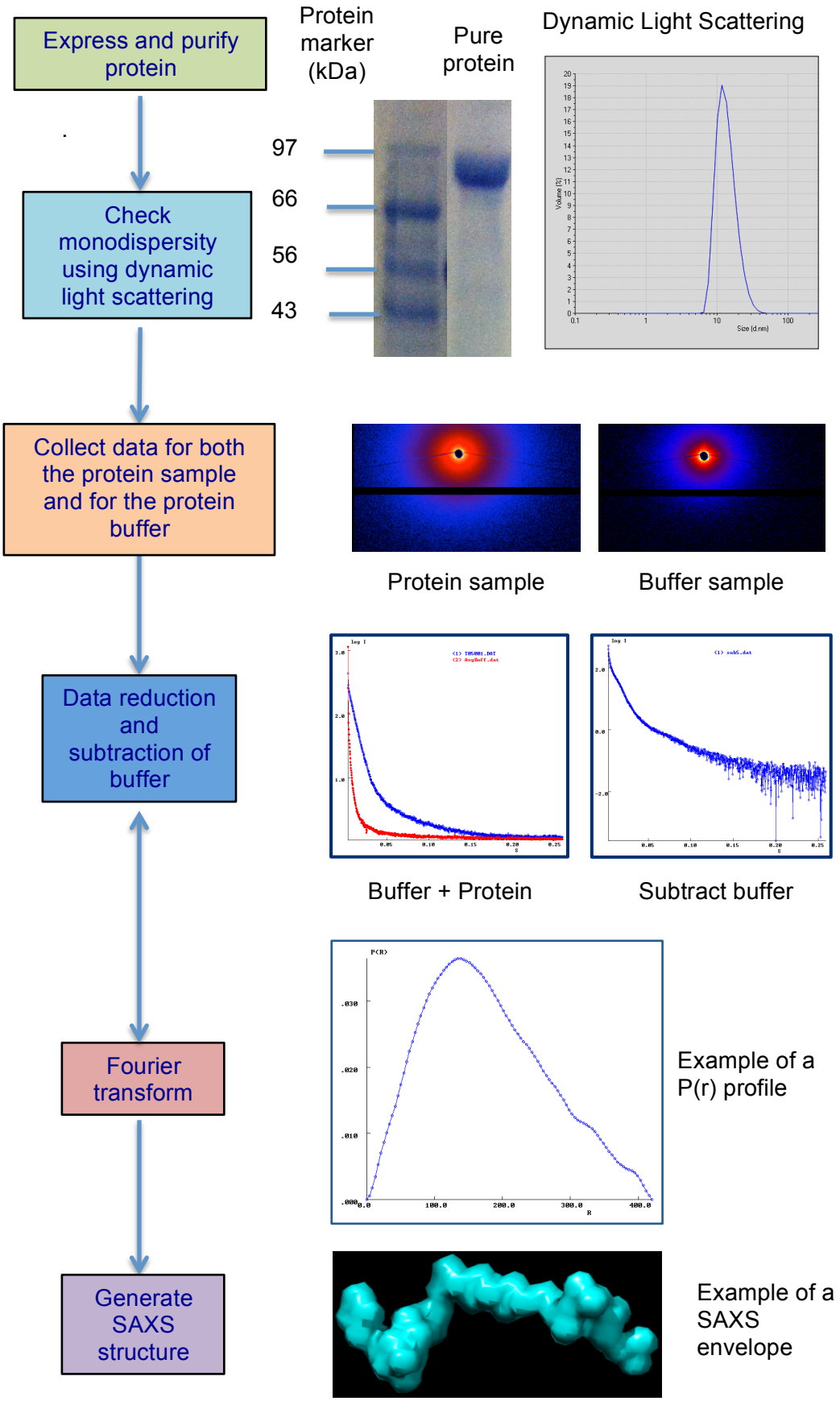


Figure 2.2.1c- Flow diagram showing steps taken to solve low-resolution structures of macromolecules using small-angle X-ray scattering

The advantage of using SAXS over crystallography is that crystalline matter is not required, which is often a limiting step in X-ray crystallography. A small sample of pure monodisperse protein is all that is required for SAXS, however there are limitations, the key restriction being that only low-resolution data is generated.

SAXS works on a similar basis to X-ray crystallography in terms of diffraction. An incoming X-ray beam illuminates the macromolecules in solution and is scattered, whilst a beam stop absorbs the direct beam. The scattered radiation is recorded on a detector, see Figure 2.2.1b.

After successful data collection, the data is reduced, and the solvent is subtracted. Following solvent subtraction, the data is Fourier transformed to obtain a radial Patterson function P(r), which describes the probable frequency of interatomic vector lengths (r) within the protein in question (Jacques and Trewhella, 2010). The P(r) profile, as seen in Figure 2.2.1c, can indicate early signs of the symmetry and domain structure within the protein. This profile can be used to calculate the scattering intensity I(0) which can not be measured directly as it cannot be distinguished from the direct beam passing through the sample unscattered. In addition the profile can be used to calculate the R_g, which is the radius of gyration, and is defined as the averaged squared distance of each scatterer from the particle centre of mass (Jacques and Trewhella, 2010; Putnam *et al.*, 2007). In the final steps of SAXS structure solving, *ab initio* algorithms are employed such as DAMMIN and DAMMIF (Franke and Svergun, 2009). Unlike crystallography, where diffraction is only obtained from quality crystals, with SAXS, scattering patterns can be collected from any quality sample. For this reason, it is crucial that the protein purity and monodispersity is analysed prior to collecting SAXS data.

The SAXS structure of Toxin B was reported in 2010 which revealed four distinct domain boundaries, into which the structural data and model structures can be aligned (Albesa-Jové *et al.*, 2010). The 4 domains are organised as demonstrated in the ABCD model in Figure 2.1.1a. The hydrophobic 'D' domain contains a highly solvent-exposed region that protrudes away from the core structure, see Figure 2.2.1d. The SAXS data were collected at a neutral pH so

the low pH induced conformational change that is predicted to occur within this 'D' domain, has not been shown in this structure.

Although the SAXS structure of Toxin B and the reconstructed electron microscopy model of Toxin A share some structurally similar characteristics, such as the elongated length feature, they do differ more than anticipated. Toxin A and B are homologous proteins, with the same domain organisation and properties, therefore some structural similarities are to be expected, but Toxin A does not appear to have the protruding hydrophobic region.

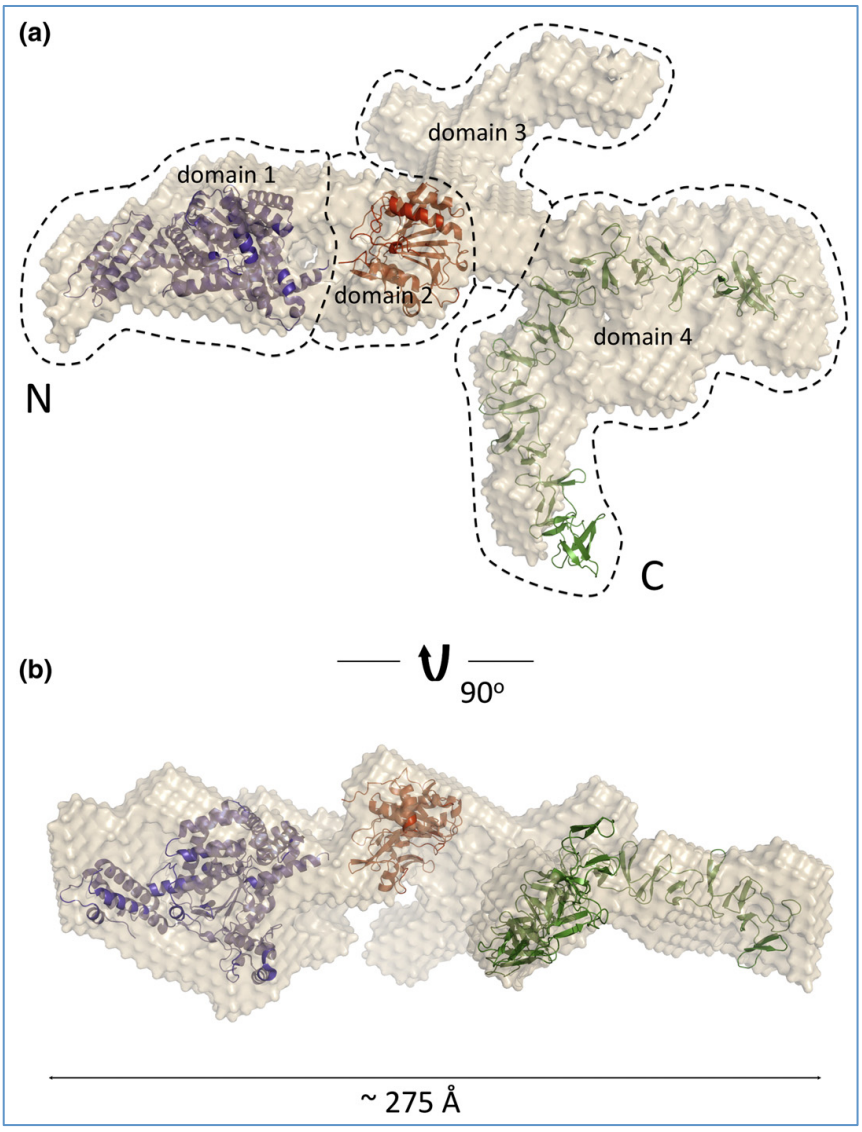


Figure 2.2.1d- SAXS structure of Toxin B. Figure adapted from Albesa-Jové *et al.*, 2010. The crystal structures of the 'A' domain, 'C' domain and 'B' domain are superimposed onto the SAXS envelope.

Dynamic Light Scattering

Dynamic Light Scattering (DLS) is a technique used to determine the size of proteins, nucleic acids, and complexes, and to detect aggregates in macromolecular solutions.

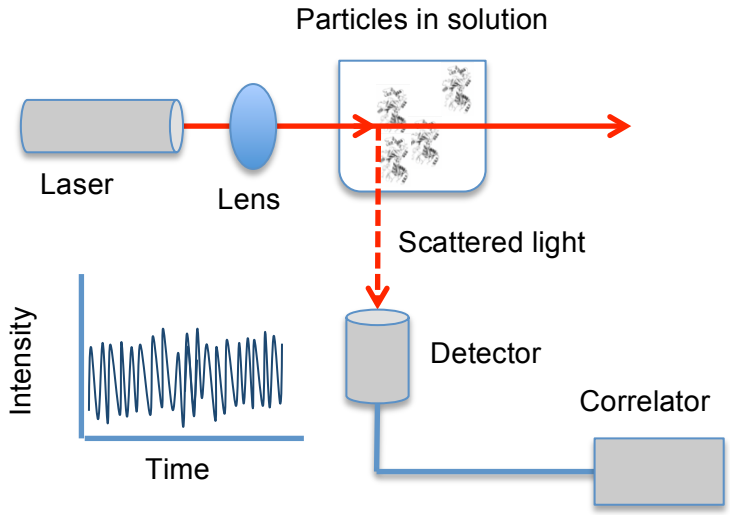


Figure 2.2.1e. A schematic diagram showing DLS data collection, adapted from Malvern Instruments.

In order to understand the concept of using DLS to determine particle size, it is important to understand that particles are moving randomly in solution by Brownian motion. A key feature of Brownian motion that is exploited in DLS is that particles of different size move in solution at different speeds, the larger the molecule the slower the movement in solution (Lorber *et al.*, 2012). The solution is exposed to a monochromatic coherent laser, through which the particles in solution scatter a small fraction of the light, the intensity of which is recorded, see Figure 2.2.1e. As the particles are moving by Brownian motion, the intensity of scattered light detected fluctuates due to the constantly changing distance between particles. This constant movement of particles causes the scattered light to undergo either constructive or destructive interference from the surrounding particles. The measured intensity fluctuations gives some information about how fast the particles are moving in solution and from that the size can be predicted. For example smaller particles have higher intensity fluctuations than larger particles as they are moving faster in solution.

The size of the particle in solution (R_H) is then calculated using the Stokes-Einstein equation:

$$R_H = kT/6\pi\eta D$$

In the Stokes-Einstein equation, k represents the Boltzmann's constant, T represents temperature, η represents the viscosity and D is the diffusion constant. The Zetasizer software autocorrelates the intensities of the scattered light for the duration of the experiment. In DLS, this time autocorrelation function of scattered light is used to determine the distribution of size across the particles in solution.

The aim of this chapter was to solve the SAXS structure of Toxin A and cross check any results with the published low-resolution toxin data collected to date.

2.2.2 Materials and Methods

Preparation of Toxin A for SAXS

Pure Toxin A was provided by PHE at 1 mg/ml concentration. This protein was concentrated to 10 mg/ml using 30 kDa cut off concentrator tubes (Amicon ® Ultra, Millipore) at 4,000 xg. The concentrated protein was then centrifuged at 13,000 rpm to remove any precipitated protein, and was loaded onto a superdex-200 gel-filtration column (GE Healthcare), as a final cleaning step. It is essential for the SAXS technique that the protein is free of contaminants and is 100 % monodisperse, hence the final cleaning step of Toxin A to ensure complete removal of contaminants and heterogeneous material. The fractions of the gel-filtration run were analysed by 8% Tris-glycine SDS-PAGE and those containing pure Toxin A were pooled and analysed by dynamic light scattering (DLS).

Analysis of Toxin A by Dynamic Light Scattering

DLS is a technique used to detect protein aggregates in solution, the size of proteins, and whether or not proteins are monodisperse in solution. DLS analysis was performed using a Zetasizer Nano S (Malvern Instruments) which uses a 4 mW He-Ne laser with wavelength (λ_o) = 633 nm, and a scattering angle of $\theta = 173°$. The protein sample was prepared by centrifugation at 13,000 rpm to remove any large particles or precipitate and was filtered through a 0.45 µm millex-HA filter (Millipore). Nitrile gloves were worn to avoid fingerprints on the sample cuvette and to prevent contamination of the sample during handling. A low volume 12 µl glass cuvette (QS 3.00 mm) was equilibrated by rinsing in the Toxin A buffer: 50 mM Tris-HCl pH 7.5 and 150 mM NaCl. A 50 µl sample of filtered Toxin A at concentration 0.5 mg/ml was loaded into the cuvette and the external surfaces of the cuvette were gently wiped with soft lens cleaning tissue to ensure no particles were present that could lead to false readings. The cuvette was placed in the sample holder and was allowed to equilibrate to 19 °C for 2 min.

A series of three scans were performed and subsequent intensity fluctuations were recorded which contain information about the time scale movement of the scattering particles. DLS results determined whether or not the Toxin A sample was suitable for use in SAXS data collection.

SAXS Data Collection and *Ab Initio* Model Building

Toxin A samples prepared as above were taken to Diamond Light Source for SAXS data collection. Experiments were performed at the I22 beamline, at 20 °C using the Pilatus 2M detector. Concentrations of 0.5 mg/ml, 1 mg/ml, 2.5 mg/ml and 5 mg/ml were used in the scattering experiments, along with four buffer samples. Calibration was carried out using BSA at 5 mg/ml in 50 mM Tris-HCl pH 7.5 as a molecular weight standard. Data was integrated using Dream (SAXS software utilities at Diamond Light Source) and the subsequent data collected for the buffer samples was averaged and subtracted from the Toxin A data using PRIMUS (Konarev *et al.*, 2003). An indirect Fourier Transformation automated by GNOM was used to generate a particle distribution function P(r) of the samples, which in turn was used to calculate the radius of gyration (R_g) and the maximum particle size (D_{max}) (Svergun, 1992). *Ab initio* modelling was performed using the on-line server for DAMMIF, whereby 20 models were produced (Franke and Svergun, 2009). The models were then averaged using DAMAVER (Volkov and Svergun, 2003). The final averaged *ab initio* model was then visualised using PyMOL and modelling of the domains in the SAXS molecular envelope was performed manually using PyMOL (version 1.5.0.4, Schrödinger, LLC).

2.2.3 Results

Preparation of Toxin A for SAXS

As it is essential when using the SAXS technique that the protein is free of contaminants and is 100 % monodisperse, a final polishing step of Toxin A was performed to ensure complete removal of contaminants and heterogeneous material, see Figure 2.2.3a. The fractions from the gel-filtration chromatography were analysed by 8% Tris-glycine SDS-PAGE and those containing pure Toxin A were pooled and analysed by DLS.

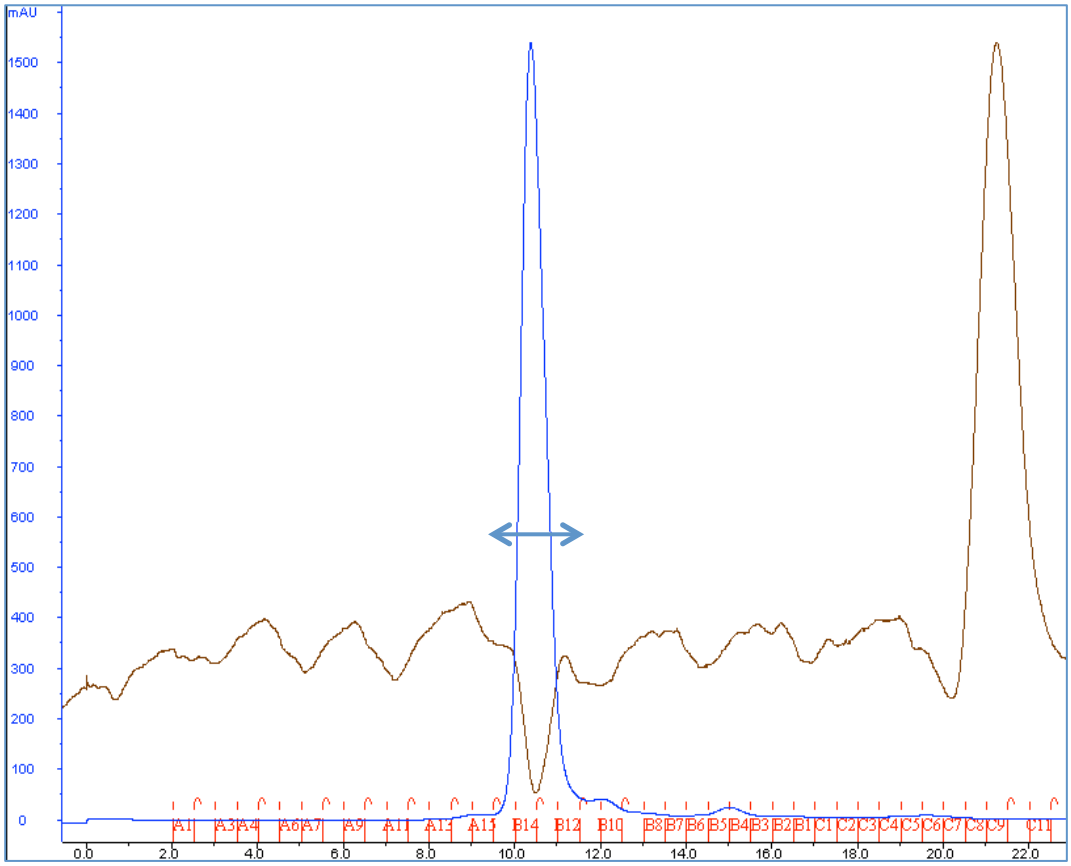


Figure 2.2.3a- Chromatogram of Toxin A size-exclusion chromatography performed on a superdex-200 10/300 GL column (GE Healthcare). The blue line represents absorbance at 280nm and the brown line represents the conductivity. Fractions A15-B12 were analysed by 8% Tris-glycine SDS-PAGE, see Figure 2.2.3b.

Figure 2.2.3b- 8 % Tris-glycine SDS-PAGE gel picture displaying fractions of the gel-filtration run containing Toxin A.

The SDS-PAGE gel picture displayed in Figure 2.2.3b shows protein bands of a molecular weight greater than 212 kDa. The gel system cannot resolve proteins of higher molecular weights than 200 kDa, hence why the protein bands are right at the top of the gel. However, it is clear that there are no contaminants of smaller molecular weight and that Toxin A is pure. To further confirm the purity of Toxin A, and to ensure that the protein is monodisperse, DLS data were collected.

Dynamic Light Scattering of Toxin A

Fractions A15-B12 containing Toxin A from the size-exclusion chromatography purification step were pooled and concentrated to 0.5 mg/ml for use in DLS experiments. Three DLS scans were performed consecutively at 19 °C for Toxin A. The Zetasizer software uses various algorithms to extract decay rates for a number of size classes to produce a size distribution.

Figure 2.2.3c-A displays the correlation function for the collected data, with the red line representing the decay over time. The decay is extended and is slow suggesting a large particle size. In addition the baseline is flat indicating a lack of larger particles or aggregates.

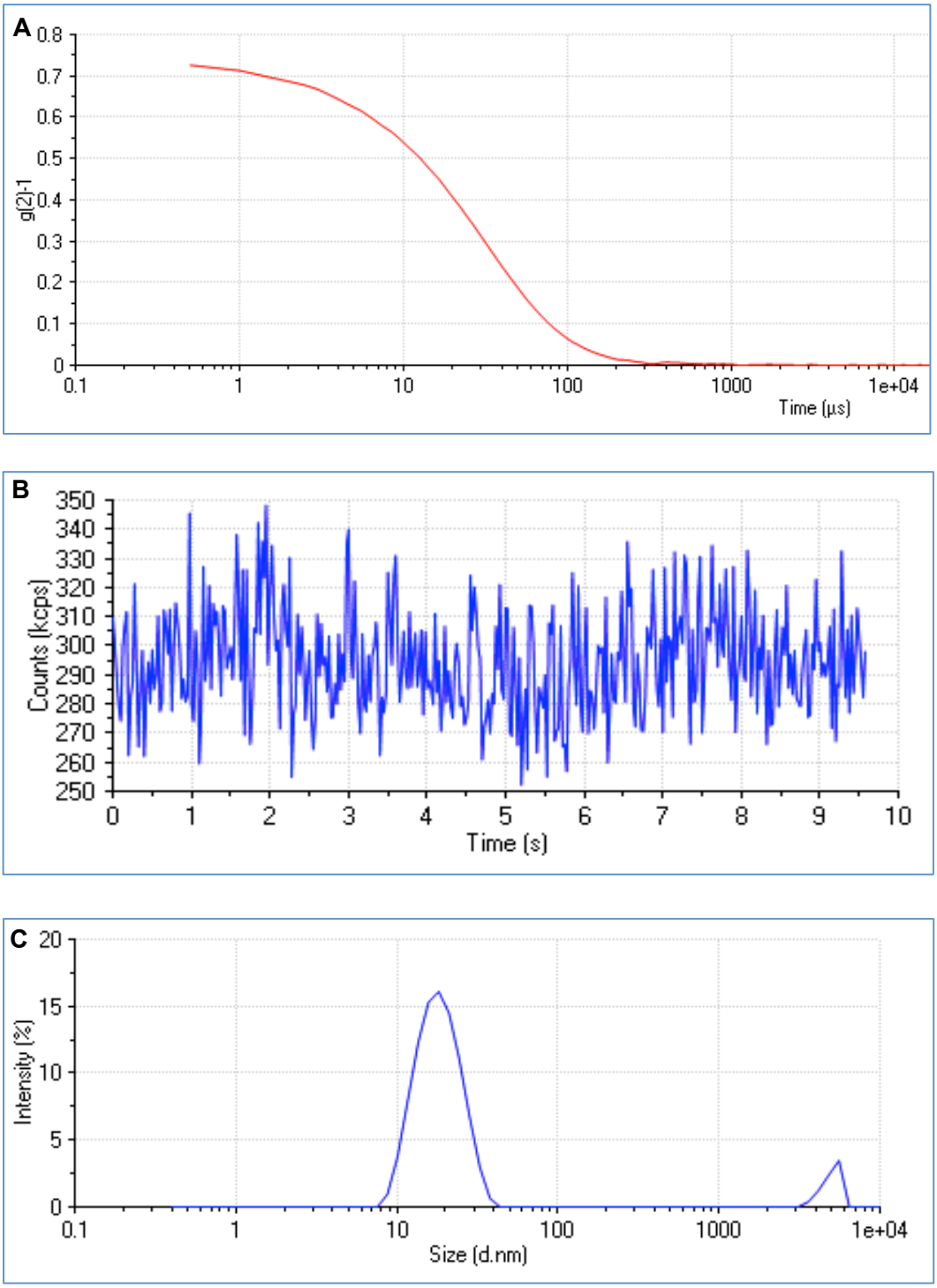


Figure 2.2.3c- DLS results for Toxin A. A: The Correlation Function showing the time at which the correlation starts to decay, providing an indication of the mean sample size. B: The Count Rate displaying the intensity fluctuations, again providing a size indicator. C: The Intensity PSD graph, which gives the intensity of each particle found and the size of the particle.

Figure 2.2.3c-B displays the count rate, which represents the intensity fluctuations and typically the smaller the particle the larger the fluctuations as the particles are moving faster in solution, whereas larger particles have lower fluctuations, such as those displayed in the figure. Finally, the intensity PSD graph displayed in Figure 2.2.3c-C provides details on the intensity of scattered light from each particle and the size of the corresponding particle. The X-axis shows the distribution of size classes whilst the Y-axis shows the relative intensity of scattered light. In this case we have a large intensity peak of about 12.5 % between 8-50 nm in diameter and one very small peak of <1 %, which is considered large and is most likely due to dust particles.

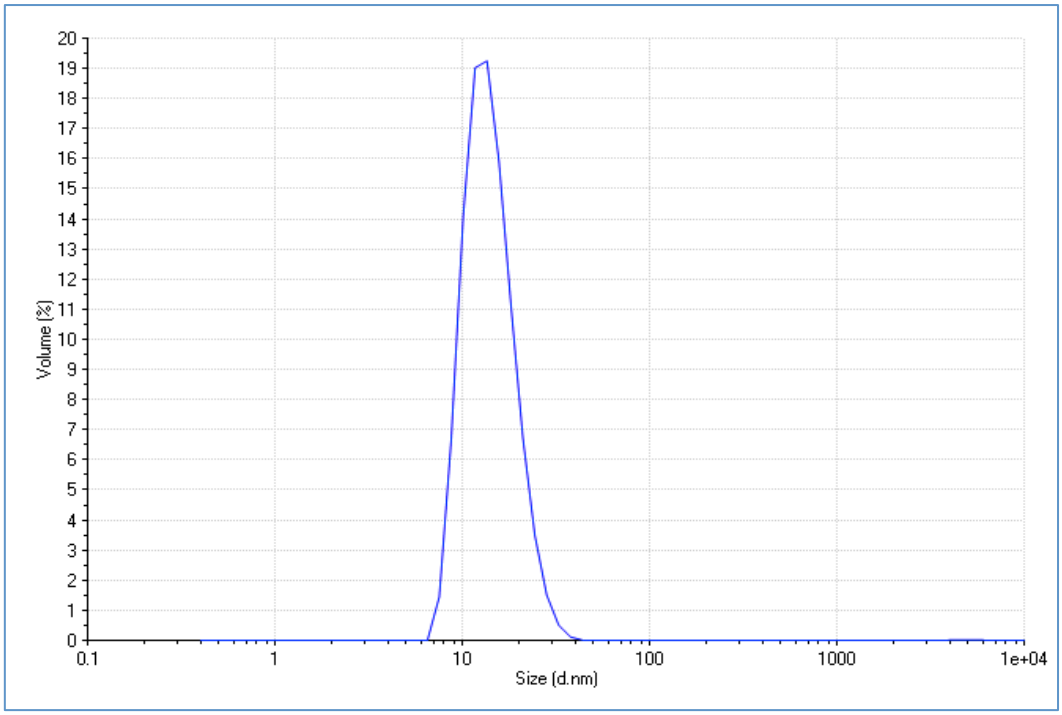


Figure 2.2.3d- Volume PSD graph for Toxin A DLS data.

The Mie theory can be used to convert the fundamental size distribution generated from an intensity distribution to a volume distribution as shown in Figure 2.2.3d (Wiscombe, 1980). Although the Mie theory assumes that the particles are all spherical, homogenous and the optical properties of the particles are known, it can be useful to examine a volume PSD graph in comparison to an intensity PSD graph. The volume PSD describes the relative proportion of multiple components in the sample based on their mass or volume (rather than their scattering intensity). From the single peak on this graph it is clear that the sample is monodisperse and suitable for SAXS data collection.

Light Scattering Data Statistics		Toxin A
Cumulants result	Z-Average size	23.36 d.nm
	PDI	0.210
Distribution result	Peak 1 (Mean/ Area)	14.39 d.nm/ 99.9 %
	Peak 2 (Mean/ Area)	4897 d.nm/ 0.1 %

Table 2.2.3a- Dynamic Light Scattering Data Statistics

The Z-Average size is also known as the cumulants mean and is defined as the hydrodynamic particle size. This measurement represents the diameter of a sphere that has the same drag coefficient as a given particle, so the Z-Average size is a predicted particle size based on the movement of the standard particles in solution. This estimated value is based on the theory that all of the particles are spherical in shape, with the Z-Average particle size diameter of 23.36 nm. As 1 nm = 10 Å we can predict that if Toxin A were spherical, it would have a diameter of approximately 234 Å. The Polydispersity Index (PDI) is a measure of the width of molecular weight distributions, and is an indication of variance in the sample. A PDI value of 0.2 or less is expected for monodisperse samples, therefore the value of 0.21 for Toxin A, is highly suggestive of monodisperse data. The distribution result describes the percentage of each identified particle in the solution along with the estimated diameter size of each of these particles, corresponding to peaks on the intensity PSD graph. Toxin A is 99.9 % monodisperse and is suitable for SAXS data collection, see Table 2.2.3a for statistics.

Small-Angle X-ray Scattering of Toxin A

Data Processing and Analysis of Scattering Data

Given the absence of structural data for the full-length Toxin A, *ab initio* modelling was performed using the program DAMMIF, whereby 20 models were generated and were averaged using the DAMAVER program (Franke and Svergun, 2009; Volkov and Svergun, 2003).

During the averaging, DAMAVER excluded 1 of the 20 *ab initio* models as it had a large normalized spacial discrepancy (NSD), therefore the final model is an average of 19 *ab initio* models, see Table 2.2.3b for detailed statistics. The distance distribution function (P(r)) from GNOM was used to calculate the radius of gyration (R_g) and the maximum particle dimension (D_{max}). The discrepancy between experimental and *ab initio* data (χ^2) was calculated by the program DAMMIF, and the NSD for the 19 models was calculated by the program DAMAVER.

Scattering Data Statistics		Toxin A
P(r) Function	R_g (Å)	89.3
	D_{max} (Å)	268.0
Structural Modelling	Discrepancy (χ^2)	1.7
	NSD	1.28 ± 0.05

Table 2.2.3b- Small Angle X-Ray Scattering Data Statistics

The R_g value was calculated using a Guinier plot, which is a form of analysis of the SAXS scattering curve at the smallest scattering angles, which corresponds to the upper portion of the scattering plot displayed in Figure 2.2.3e-A. The R_g value describes the size of the particle in terms of the root mean square distance from the outer edge of the particle to its centre. The D_{max} value is the maximum particle dimension, calculated from the P(r) curve in Figure 2.2.3e-B. The D_{max} in this case is 268 Å, which is in the same region as the Z-Average size from the DLS results. The Discrepancy (χ^2) value and NSD value are both calculated during *ab initio* modelling and averaging. The χ^2 value is a measure of how well the theoretical data fits the experimental data. Ideally the value should be less than 1.5, however this value of 1.7, is still a decent result. The NSD is a measure of similarity between the three-dimensional points of the models. NSD values above 2.0 are considered outliers and are rejected by DAMAVER for averaging. An NSD value of 1.28 suggests the models systematically differ from one another; however it is clear that the selection of models all have a similar overall shape, as indicated by Figure 2.2.3f.

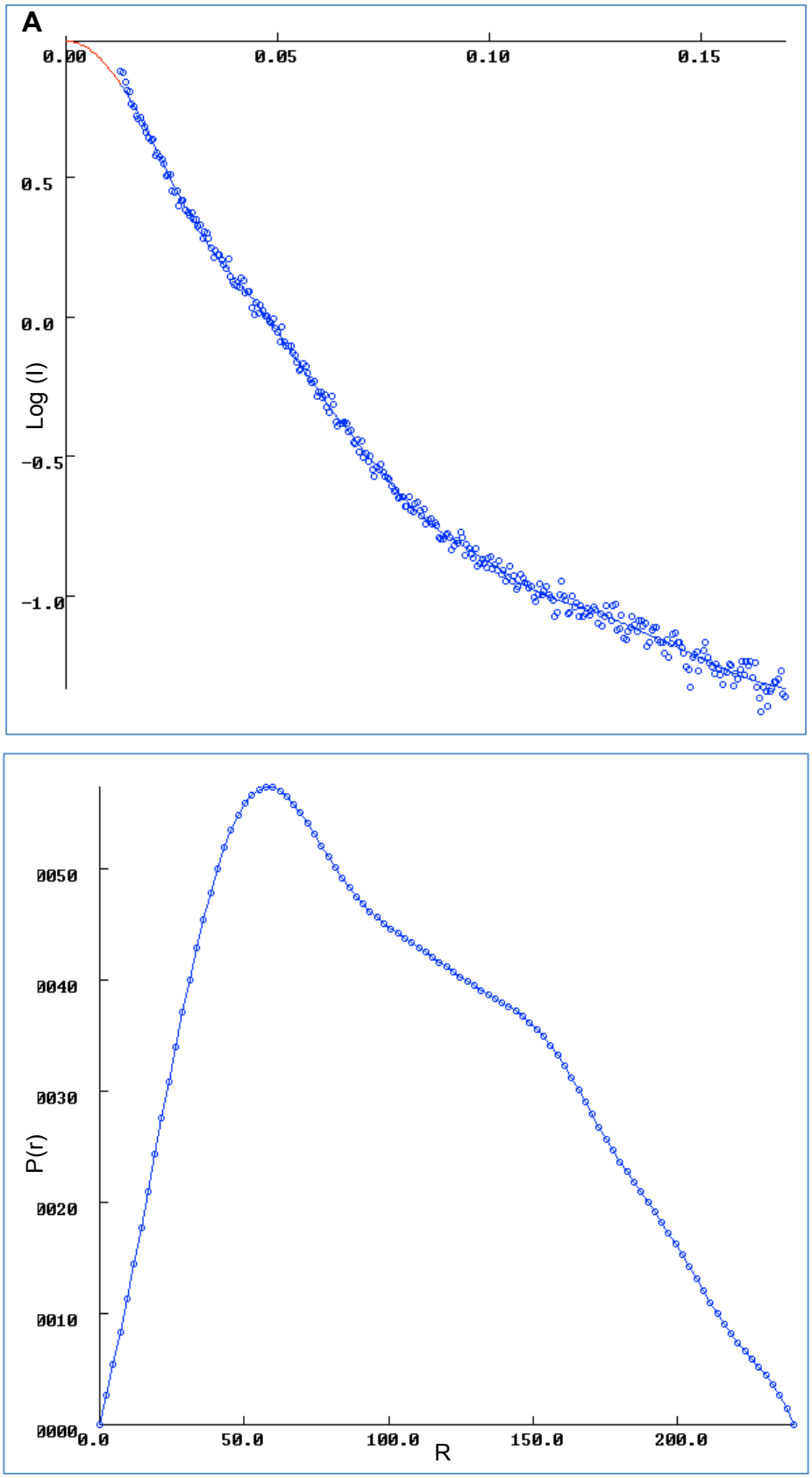


Figure 2.2.3e- SAXS scattering data. A: Scattering curve of Toxin A, where the red line represents the fit to *ab initio* data. B: The P(r) curve calculated using GNOM (Svergun, 1992).

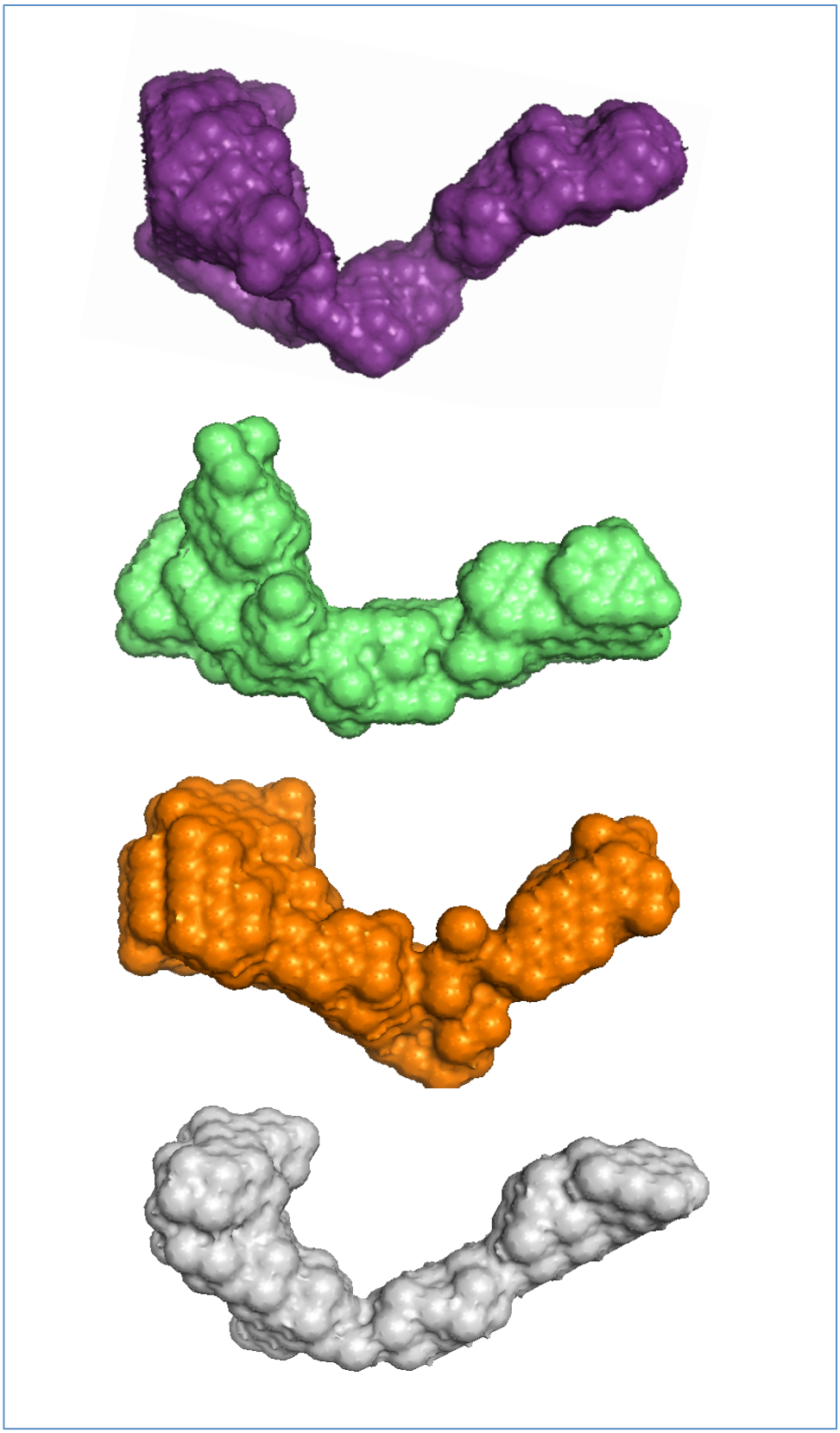


Figure 2.2.3f- Selection of four of the models generated by DAMMIF. 19 of the 20 models generated by DAMMIF were averaged using DAMAVER. Displayed in this figure are 4 of the 19 models generated, which shows the similarities between the models (Franke and Svergun, 2009; Volkov and Svergun, 2003)**.**

Ab initio Modelling of Toxin A

The final averaged *ab initio* SAXS model of Toxin A is displayed in Figure 2.2.3g.

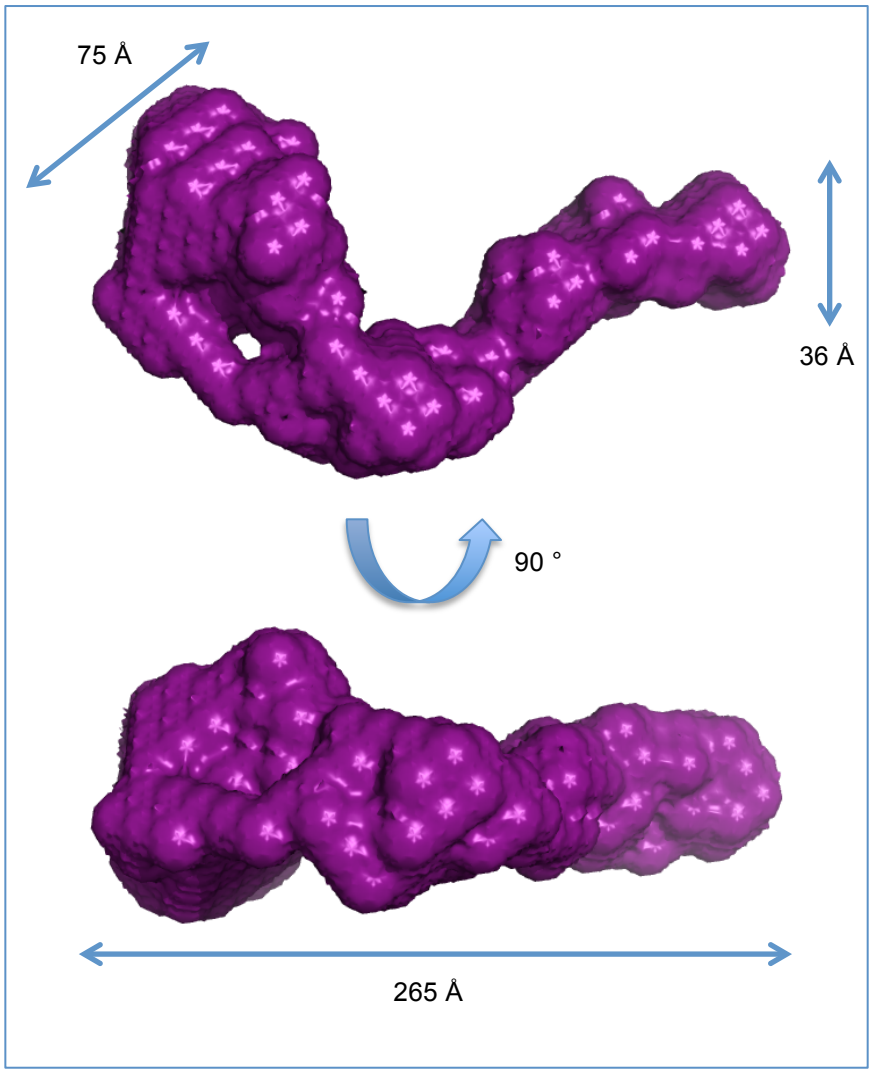


Figure 2.2.3g- The finalised SAXS structure of Toxin A after averaging 19 constructed *ab initio* models using DAMMIF and DAMAVER, displaying dimensions and a rotation along the X-axis of 90 °.

The SAXS structure of Toxin A displayed in Figure 2.2.3g, reveals an overall non-symmetrical elongated shape which is thicker and 'head-like' at one end and more narrow and 'tail-like' at the other. The structure is approximately 265 Å long, and 36 Å wide at the more narrow end and 75 Å wide at the thicker end.

Modelling of Toxin A Domains in the SAXS Envelope

The individual domains of Toxin A were placed into the SAXS envelope manually, where thought appropriate, using PyMOL (version 1.5.0.4, Schrödinger, LLC), see Figure 2.2.3b. As it is difficult to tell from the SAXS envelope where to appropriately place the domains, the domain organization as proposed by Pruitt *et al.*, with antibody labelling data, was used as an indicator (Pruitt *et al.*, 2010).

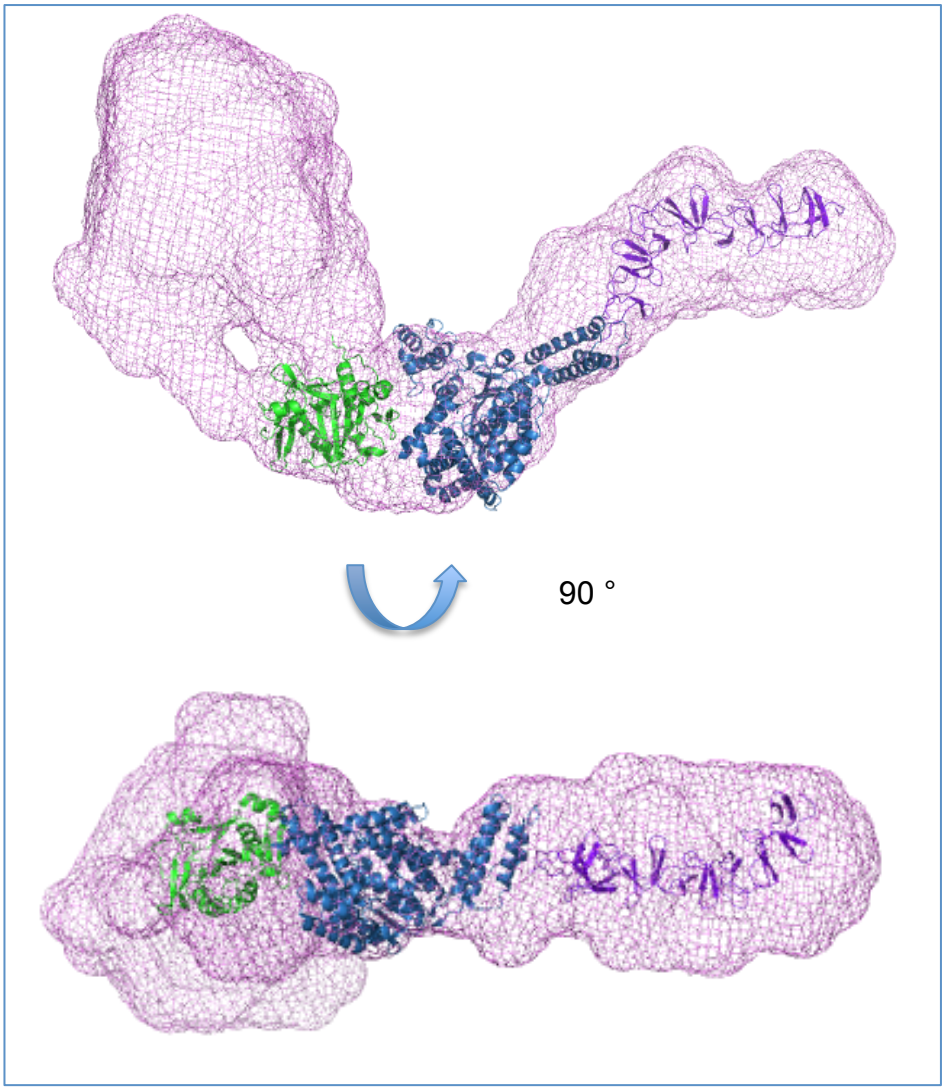


Figure 2.2.3h- Modelling of crystal structures of three of the four domains of Toxin A in the SAXS envelope. The SAXS envelope is displayed as mesh, the 'C' domain of Toxin A is displayed as a green cartoon, the 'A' domain of Toxin A is displayed as a blue cartoon, and finally the 'B' domain of Toxin A is displayed as a purple cartoon. Images rendered with PyMOL (version 1.5.0.4, Schrödinger, LLC).

The elongated tail-like portion of the SAXS envelope is thought to be the binding ‘B’ domain, which potentially extends beyond the ‘A’ domain and ‘C’ domain. The large hydrophobic region (‘D’ domain) may reside within the thicker ‘head’ region of the envelope, as this is the largest portion of the envelope unaccounted for, and the ‘D’ domain does make up for approximately half of the whole toxin.

2.3 Discussion

Crystal Growth and X-ray Crystallography

Toxin A is a large protein that shares similar properties to membrane proteins in that it has a large hydrophobic region, spanning almost 50 % of the whole toxin. It has a multi-modular domain organisation with each domain playing a key role in cytotoxicity. The 'B' domain is required for binding to target cell surfaces and the 'D' domain, which is the most hydrophobic region, is required for translocation of the toxin into the cytosol. The 'C' domain is responsible for autocatalytic cleavage of the toxin, which releases the 'A' domain into the cytosol. This 'A' domain holds the enzymatic activity which glucosylates host cell Rho GTPases causing disrupt to cellular processes including cytoskeleton maintenance (Davies *et al.*, 2011).

Toxin A contains a high percentage of leucine, isoleucine, phenylalanine and tyrosine residues, which gives the protein hydrophobic properties commonly seen in membrane proteins. The majority of the hydrophobic residues are located in the 'D' domain, as this is the membrane insertion domain. This hydrophobic element may be one of the reasons for the difficulties experienced during crystallisation. In addition, the 'B' domain structure reveals a helical region which is thought to increase surface area for optimised binding to target cells (Greco *et al.*, 2006). This region binds carbohydrate receptors on target cell surfaces. It is long, helical and may be quite flexible, which could explain why the protein molecules in the crystal are not packed tightly.

Although the Toxin A appeared recalcitrant to crystallisation during early stage crystallisation screening, eventually a crystallisation condition was obtained in which Toxin A crystals would grow but not consistently. For example, if a crystallisation plate was set up using the sitting drop method on a 96-well plate containing the exact conditions in every well, crystals would grow in approximately 3-5 of the wells, even though all of the conditions were identical. However, after extensive optimisation trials crystals that diffract up to 10 Å were successfully grown.

Attempts were made at solving the structure using the best 10 Å dataset (dataset 3), because even though the resolution is low, the data is 90.5 % complete and has an R_{symm} of 18 %. After extensive attempts at solving the structure using dataset 3 (dataset 1 and 2 were also used), as anticipated there was no success. Multiple attempts were made at solving the structure using molecular replacement, however the combination of low-resolution data and lack of homology model for 50 % of the toxin is a challenging problem. There are potentially up to 12 molecules in the asymmetric unit, so if the resolution of the crystal diffraction can be improved to at least 6 Å or higher then it may be feasible to solve the structure by molecular replacement.

Future Work Towards Obtaining Crystal Structure of Toxin A

In order to solve the structure of Toxin A, the diffraction quality of the crystals must be significantly improved. As multiple optimisation trials have been performed, the next stage in this progression should be to try co-crystallising Toxin A with various molecules to stabilise the protein. This could include carbohydrates or cell surface receptors for the 'B' domain or antibody fragments such as Fab fragments. In addition there still are a number of crystallisation methods that differ from vapour diffusion that could be tested such as microdialysis, crystallisation under oil, sandwich drop and free interface diffusion.

Low Resolution Structure of Toxin A

Due to the lack of success in high-resolution structural studies for Toxin A, low-resolution structural techniques were employed, including dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS). The DLS method provided insight into the dispersity and approximate size distribution of Toxin A. The single sharp peak on the volume PSD chromatogram indicates the protein is monodisperse and the corresponding distribution result concluded 99.9 % monodispersity. These conclusive results provide confidence that any scattering data collected during the SAXS will be that scattered by Toxin A in solution.

SAXS data collected for Toxin A were used for *ab initio* modelling to produce a low-resolution structure. The P(r) distribution function, R_g and D_{max} were calculated using GNOM, which provide information about the particle size and shape (Glatter, 1977). The P(r) curve, displayed in Figure 2.2.3e-B provides an indication as to the overall shape of the molecule in solution. A perfect bell-shape curve would normally indicate a spherical shape, whereas the curve distribution observed for Toxin A indicates an elongated and multi-modular shape (Jacques and Trewhella, 2010). The R_g value which describes the root mean square distance from the outer particle to its centre (similar to a radius) and the D_{max} value, which is the maximum particle dimension, were calculated to be 89.3 Å and 268.9 Å respectively. If the molecule was perfectly spherical, the R_g should effectively be half of the D_{max}, however in the case of Toxin A, the P(r) curve indicates an elongated structure hence why the values are not directly comparable.

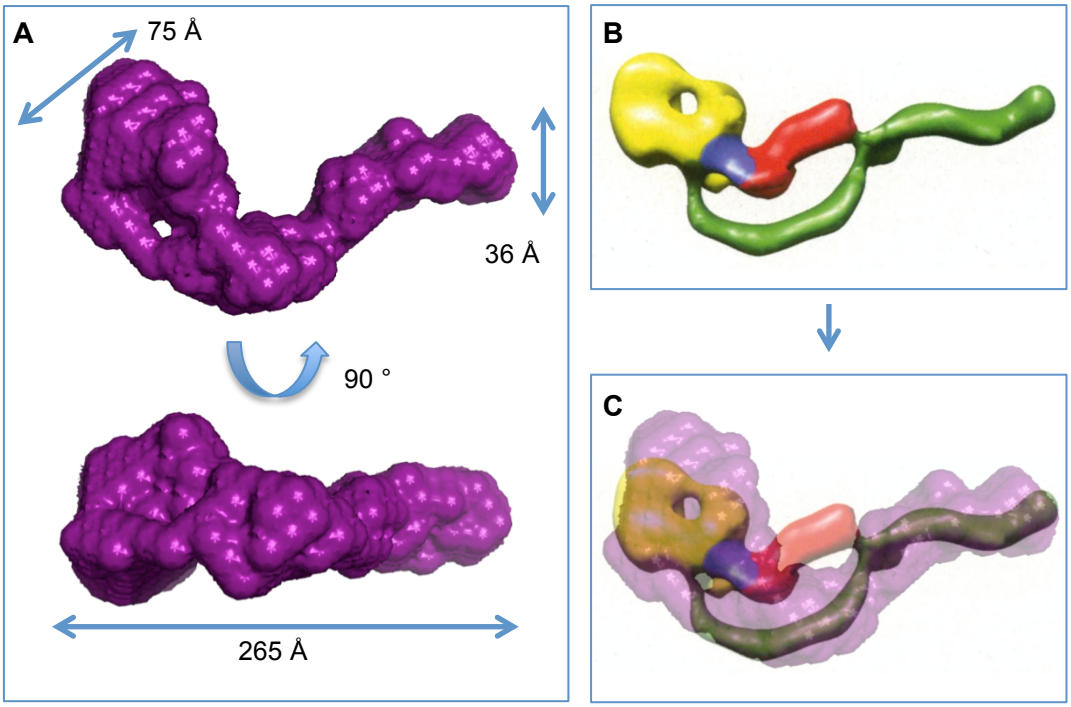


Figure 2.3a- SAXS Structure of Toxin A and comparison of the SAXS structure with the electron microscopy reconstruction. A: Finalised SAXS structure of Toxin A after averaging 19 constructed *ab initio* models using DAMMIF and DAMAVER, displaying dimensions and a rotation along the X-axis of 90°. B: 3D model structure of Toxin A constructed from electron micrographs, adapted and modified from Pruitt et al. (Pruitt *et al.*, 2010)**. C: Electron microscopy generated model of Toxin A superimposed into the SAXS envelope of Toxin A.**

DAMMIF was used to generate 20 *ab initio* models from the SAXS experimental data, which were then averaged using DAMAVER. One of the models was excluded due to the large NSD value, as calculated by DAMAVER, therefore only 19 models were used in the final averaging step to provide a final *ab initio* SAXS structure for Toxin A, displayed in Figure 2.3a-A. The structure is consistent with the statistics derived from GNOM, in that the structure is elongated rather than spherical, and the overall length is 265 Å, which is consistent with the calculated D_{max} value.

The SAXS statistics correlate with the DLS statistics in that the particle length in the SAXS structure is 265 Å and the Z-Average size particle diameter calculated from DLS data is 230 Å, which are fairly similar in size. The DLS Z-Average size particle calculation is slightly less accurate for elongated molecules, as it is based on globular proteins, hence the slight variation from that of the SAXS D_{max}.

One significant observation made during the purification of Toxin A at the gel-filtration cleaning step was that the protein eluted at a point which is to be expected for a protein of approximately 600 kDa in size, which would indicate that Toxin A is dimeric, as discussed in section 2.1.3. However, both the DLS results and the final SAXS structure indicate that Toxin A is in fact monomeric as the size distribution directly corresponds to the size observed for the low-resolution SAXS structure. This could be due to a concentration effect, as the size-exclusion step is performed when Toxin A is highly concentrated, where as DLS was performed at 0.5 mg/ml and SAXS data was collected at a low concentration range. This could be the reasoning behind the presence of dimers during purification procedures and monomers during structural data collection analysis.

The SAXS structure of Toxin A, displayed in Figure 2.3a-A shares a similar overall shape to the Toxin A reconstruction generated from electron micrographs, Figure 2.3a-B (Pruitt *et al.*, 2010). The model structure of Toxin A generated from electron microscopy data was superimposed onto the SAXS envelope of Toxin A, Figure 2.3a-C, which clearly shows their similarities. Bfrapidlyoth structures have a globular head domain and an extended long tail

domain. When comparing the Toxin A SAXS structure with the Toxin B SAXS structure, there are few similarities other than that they are both elongated and approximately 265-275 Å. The Toxin B structure lacks the globular head domain and instead has a protruding domain mid way along the envelope, see Figure 2.2.1d (Albesa-Jové *et al.*, 2010). There has not been a successful three-dimensional reconstructed structure reported for Toxin B from electron microscopy data as yet. Previous attempts found that the electron micrographs of Toxin B displayed a more heterogeneous field of particles rather than the significantly more homogeneous field of particles reported for Toxin A (Pruitt *et al.*, 2010). The characteristic globular head was still present amongst the heterogeneous particles, but due to the heterogeneity of the sample, a model was not constructed.

To conclude, elucidating the 3D-structure of Toxin A proved challenging using crystallographic techniques, however preliminary success was achieved using low-resolution structure techniques. A SAXS structure of Toxin A was obtained, which provided the first visualisation of Toxin A from X-ray scattering data. This structure was consistent with the model reconstruction structure obtained by electron microscopy. This chapter provides details of the achievements made in the contribution to characterising the structure of full length Toxin A, which needs to be explored further.

Chapter 3

CDTb

3.1 Expression and Purification of CDTb

3.1.1 Introduction

CDTb

Clostridium difficile produces two potent exotoxins, Toxin A and Toxin B, which are part of the Large Clostridial Toxin family, and induce mucosal inflammation and diarrhoea (Borriello, 1998; Kelly and LaMont, 1998; Voth and Ballard, 2005). In addition to these two exotoxins, some *C. difficile* strains produce an ADP-ribosylating binary toxin (CDT), the role of which is unknown (Perelle *et al.*, 1997; Popoff *et al.*, 1988). CDT is a member of the AB binary toxin group and so as with other members is made up of two independently produced components; the enzymatic component, CDTa (48kDa), and the transport component, CDTb (99kDa).

As discussed in Chapter 1.2, it is postulated that the enzymatic component, CDTa, irreversibly ADP-ribosylates monomeric G-actin in the host cells. The ADP-ribosylation of G-actin prevents polymerisation disrupting the G-actin–F-actin equilibrium, which causes cell rounding and cell death. The CDTa component alone is not toxic to cells, but when combined with the transport component, CDTb, cell death occurs (Sundriyal *et al.*, 2010). Therefore the aim was to characterise the transport component at a structural level in order to determine it's specific role in CDTa uptake into the cells. It is hypothesized that the CDTb component undergoes modifications including trypsin cleavage, to remove the signal peptide domain and activation domain to form CDTb''. This mature CDTb'' component is then thought to oligomerize into heptamers and interact with the target cells. Upon binding to target cells, it is thought that CDTa binds and together they are taken up into the cell by receptor-mediated endocytosis. Here, the active CDTb'' may undergo structural rearrangements upon pH changes, which occur naturally in the endosomes, inserting a pore into the endosomal membrane allowing the translocation of CDTa into the cytosol, see illustration in Figure 3.1.1a.

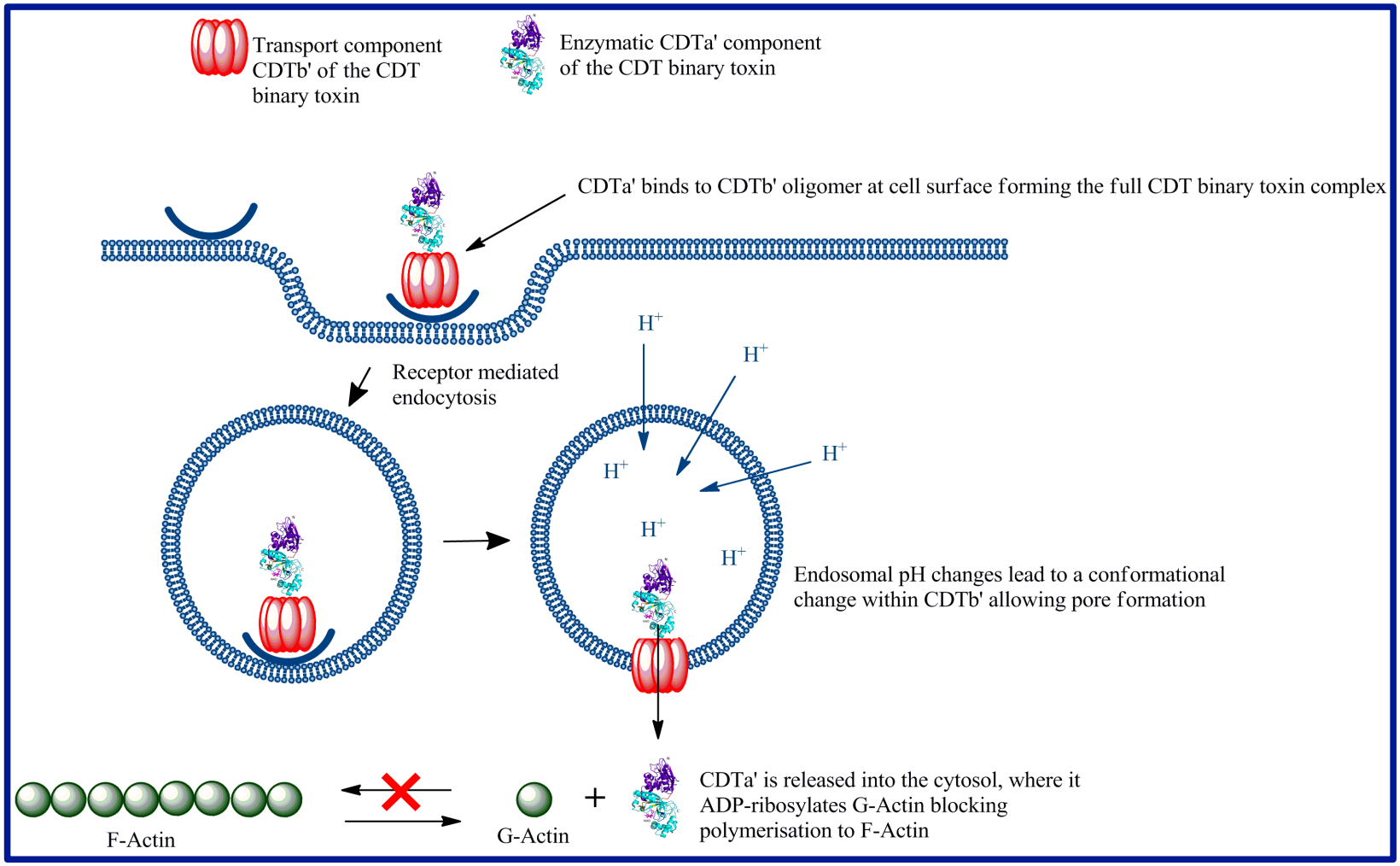


Figure 3.1.1a- Proposed mechanism of CDTa entry into target cells. **CDTa structure PDB code: 2WN7** (Sundriyal *et al.*, 2009).

In this Chapter, the aim was to characterise the CDTb component structurally, in all forms, CDTb, CDTb' and CDTb'', see Figure 3.1.1b. The prospect being that we would hopefully see the structural changes taking place during the CDTb transport process, such as the heptamer formation. In order to analyse the structural changes taking place as the CDTb becomes activated, we need to solve the structures of each of the CDTb forms, as there is no structural information available at all for any of the CDT components. The technique of choice for structure solving was X-ray crystallography. In the preliminary stages of crystallography, the first step is to obtain large quantities of protein to be purified ready for crystallisation trials.

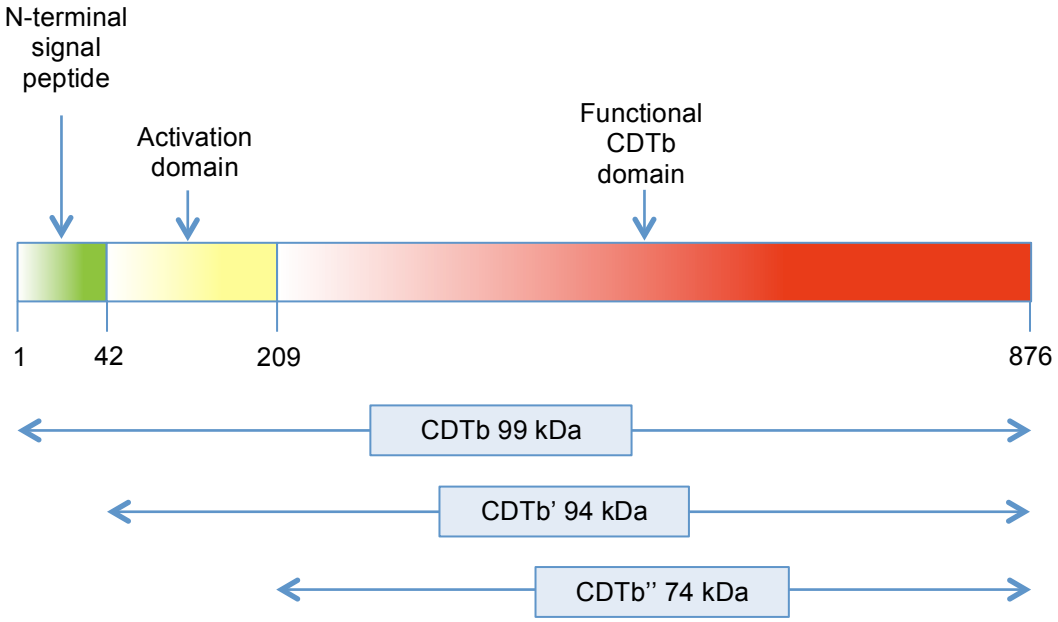


Figure 3.1.1b- Domain organisation of the CDTb variants. CDTb'' is the fully activated functional variant, CDTb' is the non-activated variant and CDTb is the full length variant with the N-terminal signal peptide and activation domain still intact.

The following section describes the process of expression of large amounts of soluble protein, followed by the purification techniques used for obtaining the required level of purity for crystallisation, for each of the CDTb forms.

3.1.2 Materials and Methods

Expression of CDTb

The previously established protocol for recombinant expression of CDTb' and CDTb'' was used as a template for expression of all variants, with some minor alterations made for optimized expression (Sundriyal *et al.*, 2010). In summary for CDTb: *E. coli* BL21 (DE3) codon plus cells (Novagen) were transformed with the pGEX-6p1-cdtB expression construct (provided by the PHE). Overexpression of soluble CDTb was obtained in a highly controlled environment using a BIOFLO-3000 bioreactor (New Brunswick Scientific) with a 4 L bioreactor culture vessel. The fermenter was prepared according to the manufacturer's manual. The pH, temperature and dissolved oxygen (DO) were controlled in a proportional-integral-derivative (PID) manner, initially programmed to pH 7.0, 37 °C, 60 % DO and agitation 200 rpm. The vessel containing 4 L of TB media (see appendix for recipe) supplemented with 1X TB salts, 0.5 % glucose and 100 µg/ml ampicillin, was allowed to equilibrate to the correct pH and temperature prior to inoculation. The medium was then inoculated with 40 ml overnight culture of the expression host, grown at 37 °C in LB (see appendix for recipe) supplemented with 100 µg/ml ampicillin. Once the culture reached OD_{600} = 0.8-1.0, the temperature was reduced to 16 °C. After the temperature reached 16 °C, 1 mM IPTG was added to induce expression of CDTb. The cultures were harvested after 20 hours, using 500 ml centrifuge tubes in a 10:500 rotor, spinning at 8,500 rpm for 10 min at 4 °C. The cell pellets were re-suspended in 2 ml/g of pellet in lysis buffer: 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 mM DTT. The re-suspended pellets were then rapidly frozen in liquid nitrogen and stored at -80 °C until further use.

The above optimised protocol for overexpression of soluble CDTb was also used for CDTb' and CDTb''.

Purification of CDTb

A protocol for CDTb' and CDTb'' purification has previously been established (Sundriyal *et al.*, 2010), however this protocol does not yield high

quantities of protein and involves a four step purification over a long period of time. This protocol serves the purpose for which it was designed. However a large yield of pure protein at high concentrations is required for structural studies. Hence the previously-established protocol required optimization. This protocol involved an initial GST-affinity step immediately followed by an anion-exchange step. The GST tag was then cleaved overnight and the following day the protein was dialyzed into an alternative buffer overnight. The next step involved another GST-affinity step followed again by an anion-exchange step. This protocol was optimized to exclude the anion-exchange step and with the addition of a final gel-filtration step.

The final optimised protocol for CDTb purification is as follows: A cell pellet, stored at -80 °C from the batches of expressed protein, was thawed and diluted in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 mM DTT) until a total volume of 10 ml/g of pellet was reached. The cells were then passed through a homogenizer twice at 20 kpsi at 4 °C. Throughout the duration of the purification from this point onwards the protein was kept on ice. The cell lysate was centrifuged at 25,000 rpm for 35 min at 4 °C and the resultant supernatant was filtered through both a mini-sart GF-pre-filter (Sartorius) and 0.45 µm millex-HA filter (Millipore). The filtered supernatant was then loaded onto a 5 ml GSTrap-FF column (GE Healthcare) at 0.5 ml/min, pre-equilibrated with lysis buffer. The column was then washed in lysis buffer until the absorbance at 280 nm reached a constant stable base line. The bound protein was eluted at 0.5 ml/min in 20 mM reduced glutathione in lysis buffer. Sufficient amount (1 U per 100 µg of protein) of PreScission protease (GE Healthcare) was added to ensure complete cleavage of the GST tag. The protein was dialyzed against 100 vol of lysis buffer, using 12-14 kDa cut off dialysis tubing (Medicell International Ltd) and was allowed to dialyze and cleave overnight at 4 °C. Dialyzed protein was then centrifuged at 4,000 g to remove any precipitated protein and was re-loaded back onto the GSTrap-FF column, pre-equilibrated in lysis buffer, at 0.5 ml/min, collecting the pure CDTb in the flow through. The column was then washed until a constant stable base line was reached and the bound GST tag was eluted using 20 mM reduced glutathione in lysis buffer. If any GST tag was still present after the above 2-day purification protocol, the protein was concentrated to 1 ml and loaded onto a superdex-200 10/300 GL column (GE Healthcare) to achieve complete separation. The final pure CDTb was concentrated to 3.9 mg/ml using

30 kDa cut-off hydrosart membrane concentrators (Sartorius, Vivaspin), rapid frozen in liquid nitrogen and stored at -80 °C.

The above optimised purification protocol was also used for the production of high yields of pure CDTb' and CDTb''.

Verification of CDTb Variants

In order to confirm the pure proteins are in fact the CDTb variants, attempts were made at analysing the samples by Mass Spectrometry (Mass Spec). CDTb, CDTb' and CDTb'' were all concentrated to 10 mg/ml and 0.1 % formic acid was added in preparation for Mass Spec. For reasons that will be explained in the results section, another batch was concentrated and 0.1 % acetic acid was added. In addition, a separate batch of proteins were slowly dialysed into 0.1 % formic acid and concurrently 0.1 % acetic acid.

A Western blot was performed using an antibody specific for GST, to confirm the presence of GST on the blot, particularly before and after cleavage of the GST-tag from CDTb to ensure that the proteins being expressed and purified are in fact CDTb variants.

3.1.3 Results

Expression of All CDTb Variants

The expression of all three CDTb variants was optimised to provide a high yield of soluble protein for use in purification trials. CDTb variants were all expressed using the same optimised protocol whereby *E. coli* BL21 (DE3) codon plus cells were transformed with the pGEX-6p1-cdtB expression constructs (provided by the PHE). Over-expression of soluble CDTb variants was obtained in a highly controlled environment using a BIOFLO-3000 bioreactor (New Brunswick Scientific) with a 4 L bioreactor culture vessel. Samples of the cultures were taken prior to induction with IPTG and post induction. Expression levels were analysed by 10% Bis-Tris SDS-PAGE, see Figure 3.1.3a.

Figure 3.1.3a displays the three gel pictures from the SDS-PAGE with both the pre-induced and post-induced samples in lanes 2 and 3 respectively, and the markers in lane 1. For the CDTb variant in gel picture A, it is clear that there is no protein at approximate molecular weight 127 kDa prior to induction, and then a significantly large band at post-induction, indicating successful expression levels. CDTb' also shows good expression levels as indicated by gel picture B, where there is little sign of the protein prior to induction and then a larger protein band at approximately 122 kDa after induction. CDTb'' again shows successful expression levels post-induction in gel picture C at 103 kDa. The comparison between pre-induced and post-induced samples provides an indication that the cells are expressing a protein that is not normally expressed upon addition of IPTG.

For proteins CDTb and CDTb' the expression bands on the gel are slightly lower than expected for proteins of molecular weight 127 kDa and 122 kDa respectively, which is most likely due to overloading of protein on the gel.

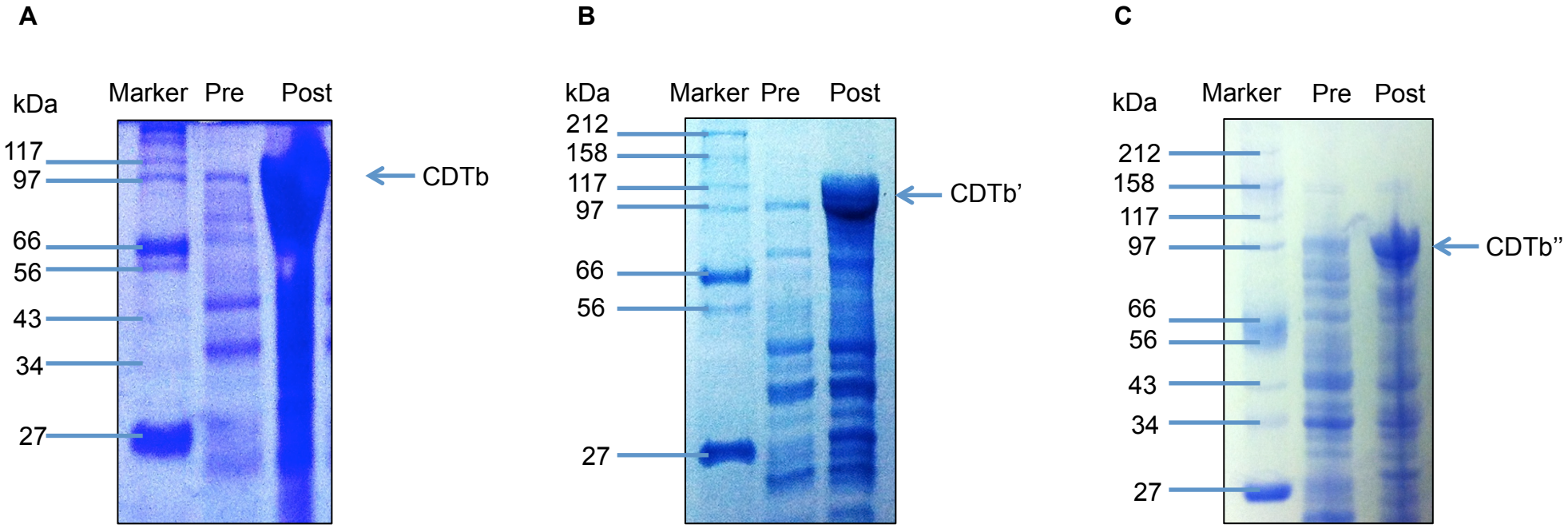


Figure 3.1.3a- Expression of CDTb, CDTb' and CDTb'' analysed by 10 % Bis-Tris SDS-PAGE. **Marker displayed on all images is the broad range 2-212 kDa protein parker (New England BioLabs ®) A: Expression of CDTb shown by the presence of a strong band at molecular weight 100 kDa that is not present in the pre-induced samples. B: Expression of CDTb' shown by the presence of a strong band at molecular weight 100 kDa that is not present in the pre-induced samples. C: Expression of CDTb'' again shown by a strong band at molecule weight 100 kDa, not observed in the pre-induced sample.**

Purification of CDTb Variants

A protocol for the purification of CDTb' and CDTb'' had previously been established however the resultant output from this protocol is a low yield of pure CDTb' and CDTb'' after four days. Therefore, the protocol was optimised to provide an overall higher yield during a shorter time frame. The final optimised protocol for all CDTb variants involved a 3-day protocol, in brief, as follows:

Day 1:

- Cell lysis using cell disruptor
- Centrifugation to remove cell debris
- Supernatant preparation including filtration
- GST-affinity chromatography
- Cleavage of GST-tag using PreScission Protease
- Dialysis overnight to remove reduced gluthatione

Day 2:

- Spin cleaved protein to remove any precipitate
- GST-affinity chromatography to remove tag
- Collecting pure CDTb protein in the flow through
- Concentration pure CDTb protein
- Size-exclusion chromatography overnight

Day 3:

- Analyse size-exclusion fractions
- Pool and concentrate pure fractions of CDTb protein
- Perform structural or functional experiments

As observed in the chromatogram for Day 1 of the purification step, Figure 3.1.3b-A, the signature profile for GST-affinity chromatography initially starts with a large broad peak, which represents non-GST-tagged proteins that do not bind to the GSTrap column. During this time, GST-tagged CDTb protein binds the GSTrap column. Once the supernatant was loaded, the column was washed in buffer until a steady flat UV baseline is reached. Reduced glutathione was then added to the buffer to elute the GST-tagged CDTb protein from the column. Once the UV reading reached a steady flat base line again, the elution was terminated.

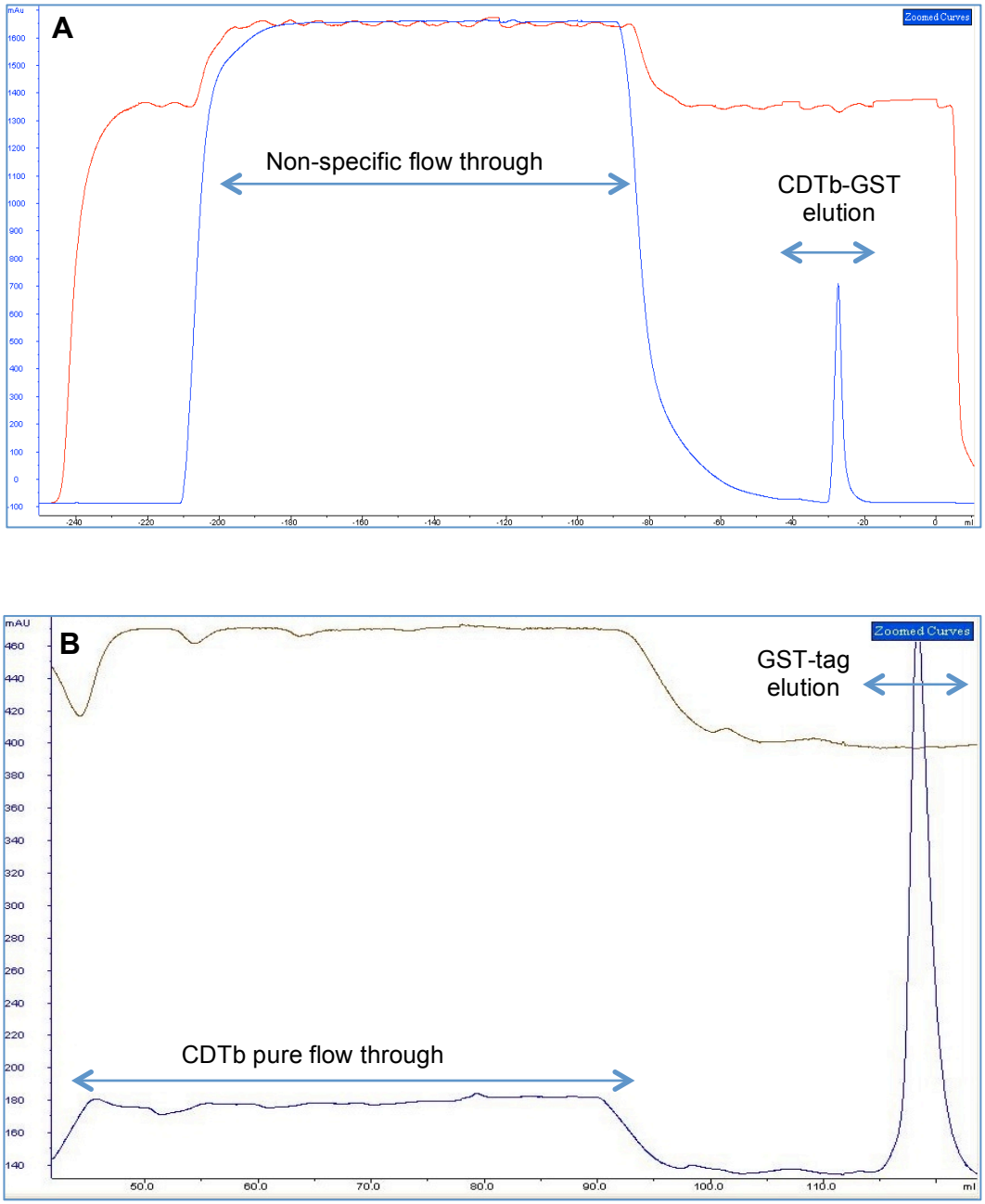


Figure 3.1.3b- Purification of CDTb variants. **The blue lines represent the absorbance at 280 nm and the brown/red lines represent the conductivity. A: First step of GST-affinity chromatography with GST-tagged CDTb binding to the GSTrap column. Non-GST-specific proteins do not bind the column and pass through in the flow through. The CDTb-GST is eluted from the column using reduced glutathione. B: The second step of GST-affinity chromatography. After the eluted protein from step 1 is treated with PreScission Protease, the CDTb does not bind the GSTrap but is collected in the flow through. The bound GST-tag is then eluted in reduced glutathione as previously.**

The collected CDTb-GST elution was dialysed to remove the reduced glutathione and PreScission protease was added to cleave off the GST-tag. In the second step GST-affinity chromatography CDTb would no longer bind to the GSTrap column and was collected in the flow through, see Figure 3.1.3b-B. The GST-tag was then eluted using reduced glutathione. Occasionally this method doesn't allow complete removal of the GST-tag, therefore a final cleaning step by size-exclusion chromatography can be performed, see Figure 3.1.3c.

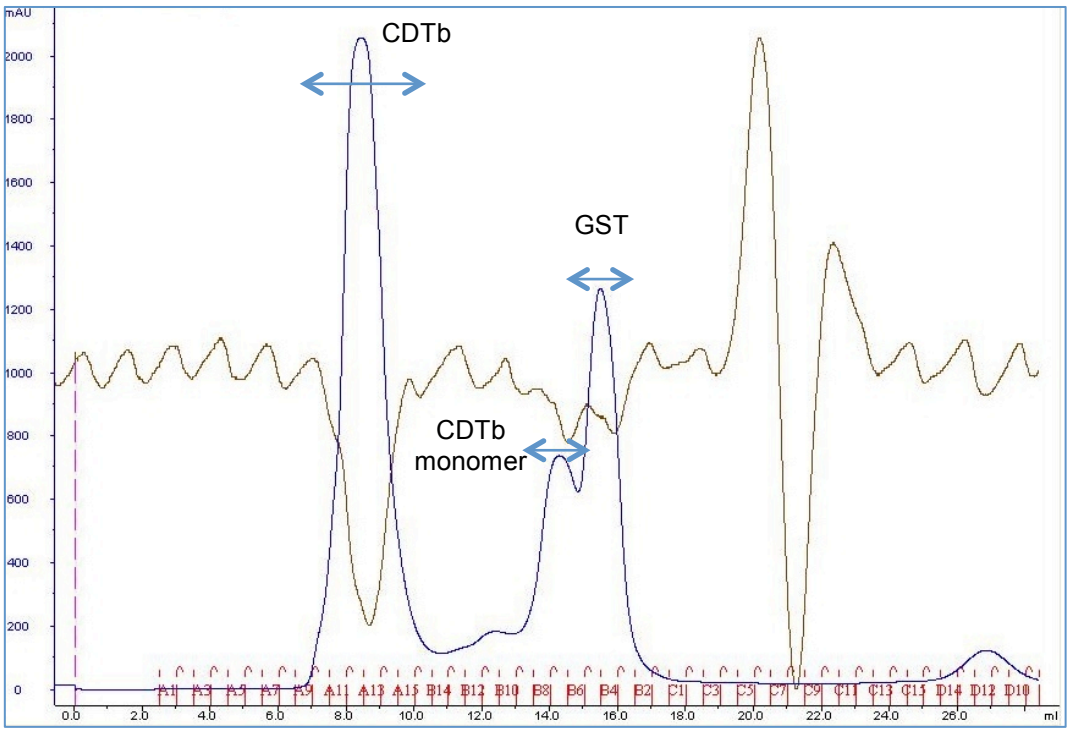


Figure 3.1.3c- Example of gel-filtration chromatography step for all CDTb variants to ensure complete removal of GST-tag. The blue line represents the absorbance at 280 nm and the brown line represents conductivity. All of the CDTb variants form oligomers that start to elute from the superdex-200 10/300 GL at approximately 7 ml. A monomeric peak is often visible at approximately 14 ml if full oligomerisation has not taken place. Finally the GST-tag elutes at approximately 15-16 ml.

All of the CDTb variants are purified using the same protocol, and each behaves in a similar manner throughout the purification. The first two steps using GST-affinity chromatography were identical as expected, however all three CDTb variants elute at the same point which is predicted for proteins greater than 500 kDa. This was unexpected as it is thought that only the CDTb'' which has been trypsin activated forms oligomers. In addition, the predicted size is larger than the calculated heptamer size.

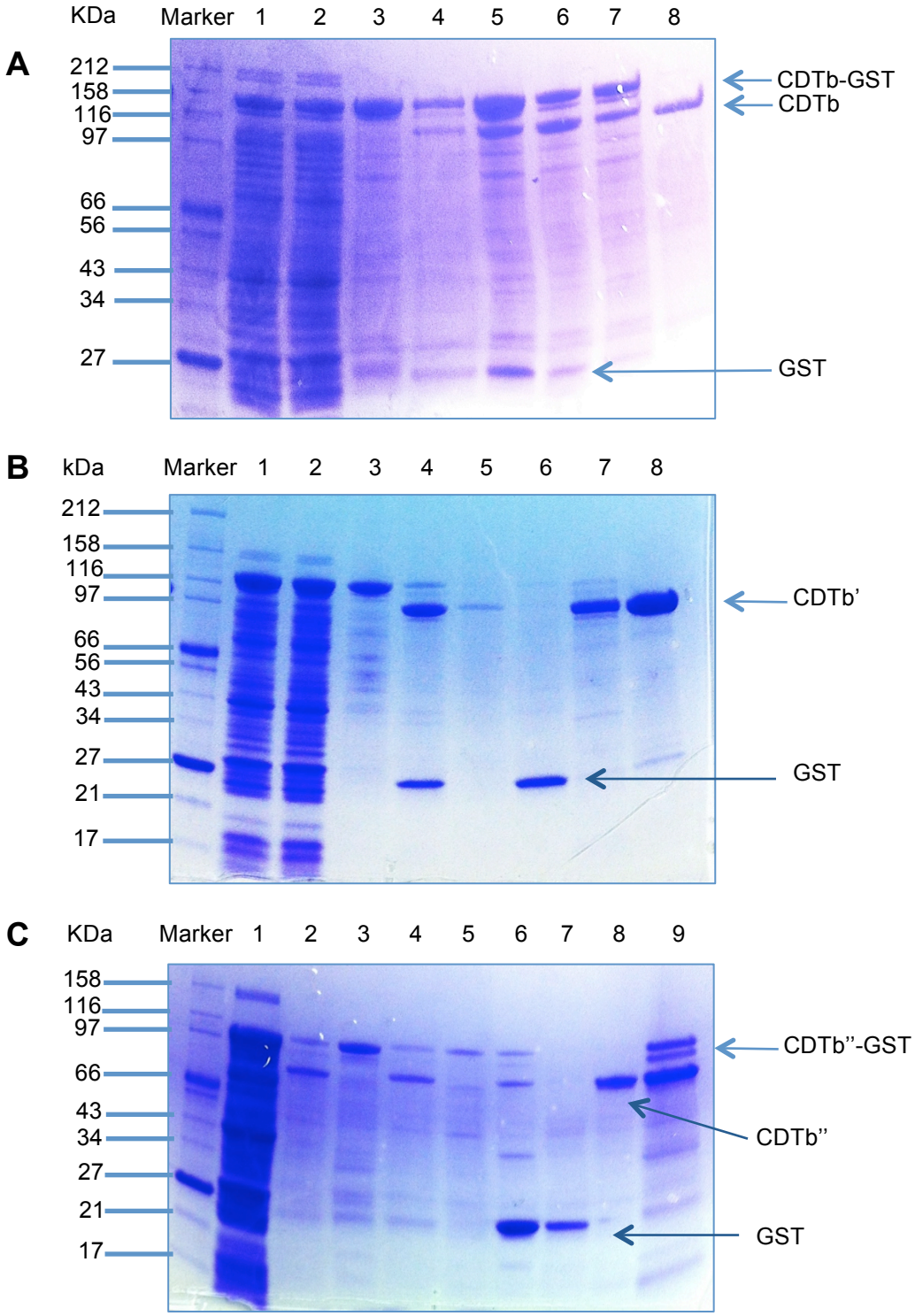


Figure 3.1.3d- 4-15 % Tris-glycine SDS-PAGE gel pictures displaying purification steps for each CDTb variant. **For each image lanes 1 and 2 correspond to the protein loaded onto the column and the subsequent flow through respectively. Lanes 3 and 4 represent the eluted GST-tagged protein and the dialysed cleaved protein loaded back onto the GST column respectively. Lanes 5 and 6 represent the 'pure' protein flow through and the eluted GST-tag respectively. For gels A and B, lanes 7 and 8 represent peaks 1 and 2 from the size-exclusion chromatogram, whereas for gel C lanes 7, 8 and 9 represent the third, second and first peaks respectively. A: CDTb purification displaying final product in lane 8. B: CDTb' purification displaying final pure product in lane 8. C: CDTb'' purification displaying final pure product in lane 8.**

Figure 3.1.3d displays the CDTb variants at each step during the purification protocol as analysed by 4-15 % Tris-glycine SDS-PAGE. For gel pictures A, B and C the first 4 lanes represent the cell supernatant load, the non-specific flow through, the CDTb-GST elution and the PreScission protease cleave result respectively. For all three of the variants, it is clear from lane 4 that the GST-tag has been cleaved, as the CDTb-GST band from the elution step in lane 3 has dropped down in size by approximately 30 kDa in lane 4, and a band at 30 kDa has appeared representing the cleaved GST-tag. However, in some cases, particularly obvious with CDTb (gel A) there is not complete cleavage suggesting that sufficient amounts of PreScission protease were not added. However, step 2 in the purification using the GSTrap column removes most of this uncleaved protein and any remaining is removed during size exclusion. Lane 8 on all three gels displays the final pure product for all three variants, with each protein showing > 95% purity.

The unusual observation noted during the purification of CDTb variants was in the final cleaning step of gel-filtration chromatography. All of the variants eluted earlier than is expected for proteins of their size. In addition, if the fully processed CDTb'' variant were to form heptamers as predicted, the protein would certainly elute earlier, however, all of the variants are eluted from the size-exclusion column at a point which is to be expected for proteins greater than 500 kDa. Brief attempts were made at trying to prevent oligomerisation for CDTb'' whereby the large oligomeric elution peak was passed back through the size-exclusion column pre-equilibrated with a range of different buffers targeted towards preventing oligomerisation. Such buffers included the standard 50 mM Tris-HCl pH 7.5 with increased salt (200 mM NaCl, 300 mM NaCl and 500 mM NaCl), increased DTT (2 mM, 5 mM and 10 mM DTT) and addition of detergent Tween-20 (0.1 %, 0.2 %, 1 % and 2 %). As a result there was no shift in the elution peak, CDTb'' continued to elute at the same point.

Verification of CDTb Variants

Multiple attempts were made to prepare CDTb pure proteins for Mass Spectrometry, however addition of acid using any of the methods described in Chapter 3.1.2 resulted in complete precipitation of all CDTb variants.

As Mass Spectrometry was not a suitable option for CDTb, CDTb' and CDTb'', as the proteins crash out of solution upon addition of acid, a Western blot was performed using a GST-specific antibody for the CDTb' variant. The samples analysed on the gel were those taken during the various steps of purification, therefore GST should appear at approximately 130 kDa when bound to CDTb'' and at 27 kDa after cleavage with PreScission protease, see Figure 3.1.3e.

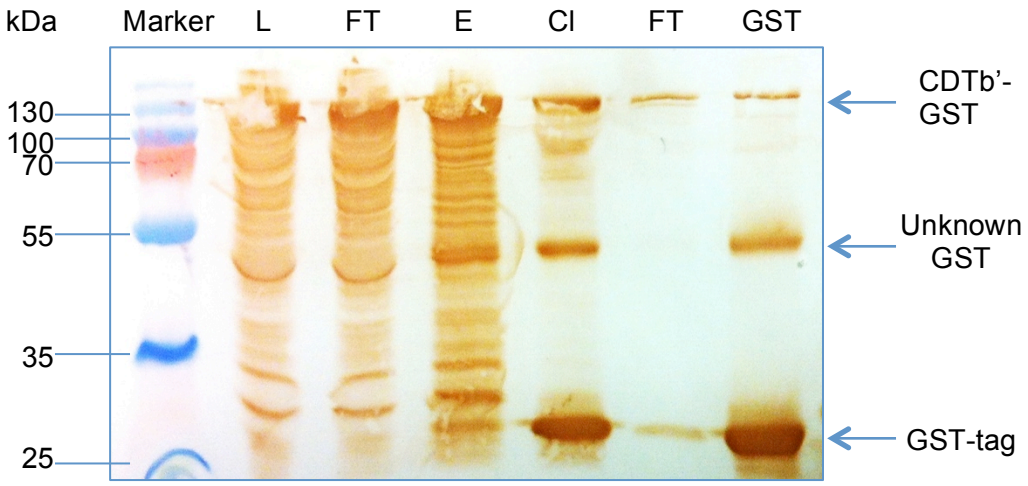


Figure 3.1.3e- Western blot using GST-specific antibody to confirm purification of CDTb'. 'L' and 'FT' lanes represent the loaded cell supernatant and non-specific flow-through from GSTrap column. 'E' represents the CDTb'GST elution from the column. 'Cl' represents PreScission protease cleavage. 'FT' represents the flow-through when passed back through a GSTrap column after cleavage. 'GST' represents the final elution from the GSTrap column using reduced glutathione.

The Western blot displayed in Figure 3.1.3e represent the first two steps in the purification of CDTb'. CDTb'-GST bands are present in the load, flow through and elution fractions as observed in lanes 'L', 'FT' and 'E' on the blot, at approximately 130 kDa. There are numerous other bands present, but at this point in the purification there are still multiple *E.coli* proteins present. After cleavage, the band at 130 kDa decreases in size by approximately 30 kDa and a strong band appears at 27 kDa indicating GST cleavage from CDTb'. This is a strong indication that the correct protein is being expressed and purified, as the PreScission protease cleavage site is highly specific, and the bands are of the correct molecular weight. The 'FT' and 'GST' lanes representing the CDTb' pure

flow through (with some small CDTb'GST and GST contamination as normally observed at this point) and the final GST elution are as expected. There is still some uncleaved CDTb'-GST present across all of the lanes, which is not uncommon during purification. There is however an unusual band at approximately 40 kDa which appears prominent at the elution step and is present throughout the purification. It is GST specific, hence it is flagged up in the Western blot, but it clearly doesn't affect CDTb' purification as it is not present in the second flow-through step which contains pure CDTb'. It has a high binding affinity for the GST-column, but is not overexpressed as it does not stand out as a strong band in the 'L' and 'FT' lanes. It is likely to be an *E.coli* protein with GST properties.

3.2 Crystallisation of CDTb

3.2.1 Introduction

CDTb

Clostridium difficile produces an ADP-ribosylating binary toxin (CDT) that is made up of two independently produced components: CDTa and CDTb. As mentioned in the introduction, the role of CDT in pathogenesis is unknown, however what is understood is that both components are required for toxicity (Sundriyal *et al.*, 2010). As discussed previously in Chapter 3.1.1, there are three forms of CDTb; the full length early form (CDTb) which has both an N-terminal signal peptide domain and an activation domain, the pre-activated form which does not have the signal peptide domain (CDTb') and lastly the active pore-forming form without the activation or signal peptide domains (CDTb''), see Figure 3.1.1a.

The aim of this subchapter was to crystallise the varying forms of CDTb and solve their structures using X-ray crystallography, in the hope that this structural information would contribute to the understanding of CDTb processing from the inactivated form to the mature oligomeric form.

Crystallisation

Crystallisation techniques employed in this chapter have been described extensively in Chapter 2.1.1.

As will become clear in the following sections of this chapter, crystallography can be a challenging field of biology, but some preliminary results show a promising outlook for further research into CDTb structural analysis.

3.2.2 Materials and Methods

Crystallisation of CDTb variants

Preliminary Screening

High yields of CDTb, CDTb' and CDTb'' were obtained using the optimised expression and purification protocols in Chapter 3.1.2. A pre-crystallization test (PCT) was performed to determine a suitable concentration range for optimal crystal growth, see Chapter 2.1.2 for details on PCT protocol.

A 4-well crystallisation plate was set up using the four reagents described in Table 2.1.2a. The PCT was set up using the hanging drop vapour diffusion method with a drop size of 1:1 µl ratio of protein to reservoir. The PCT plate was incubated at 16 °C for 30 min before being analysed using a light microscope. If the concentration of the protein was deemed suitable for crystallisation in the PCT, then the preliminary screens were performed. Unless stated otherwise, preliminary screens were performed with the Phoenix Liquid Handling System (Art Robbins Instruments) using the sitting drop method with a ratio of 0.3:0.3 µl protein to reservoir ratio.

The following 96-well crystallisation screens were used: Clear Strategy Screen I HT-96, Clear Strategy Screen II HT-96, JCSG-plus HT-96, Heavy + Light twin pack HT-96, Morpheus HT-96, MIDAS HT-96, MemGold HT-96, PGA screen HT-96, ProPlex HT-96, PACT premier HT-96, Stura Footprint Combination screen HT-96, and finally Structure screen I + II HT-96 (All screens were purchased from Molecular Dimensions). The crystallisation screens were all incubated at 16 °C, and monitored initially after 24 hours and then fortnightly thereafter. The concentration at which the above screens were performed varies for each of the CDTb forms, based on the PCT results, and will be discussed in more detail in Chapter 3.2.3. In such cases where no hits were observed during preliminary screening, the above screens were also performed at 21 °C and 4 °C, and at a variety of concentrations, to encourage crystal growth.

Optimisation of Preliminary Hits

After observation of any initial hits during the preliminary screens, the corresponding crystallisation conditions were further tested to optimise crystal growth. Optimisations were performed on 24-well plates using either the hanging drop or sitting drop method, with a reservoir volume of either 0.5 ml or 1.0 ml. Generally speaking, crystallisation screens have various combinations of buffers, salts and precipitants that promote crystal growth. In order to determine the optimal condition, the individual components are varied independently. Typically the precipitant concentration was varied first at between 0-30 %, and in some cases the precipitant was substituted for an alternative of different molecular weight. The salt concentration was varied to ±10-100 mM of the original concentration, and was in some cases substituted with other salts containing similar elements. The buffer pH was also varied by up to two units on either side of the original pH, however this buffer was rarely substituted for an alternative. Temperature and method of vapour diffusion were also variables tested for optimisation.

Alternative Techniques to Promote or Optimise Crystal Growth

Under such circumstances where no initial hits were observed, or crystal hits showed little or no diffraction, further steps were taken to encourage crystal growth. One such step involved buffer exchanging the proteins into a different buffer and pH, or alternatively water, prior to the initial screens.

In some cases where initial hits were observed, but the crystals showed no diffraction, or in cases where optimisations had no effect on crystal quality, additional additive screens were performed. When performing additive screens, the initial condition was used as the primary reservoir and then various compounds were added in small quantities. The additive screens performed include Silver Bullets and Additives Screen HT (Both from Hampton Research).

For some of the CDTb forms, seeding technique were used in an attempt to improve crystal quality and size. Further details and results will be provided for each of the techniques used in Chapter 3.2.3.

3.2.3 Results

Crystallisation of CDTb

Preliminary Screens

Preliminary screens for CDTb were set up as discussed in Chapter 3.2.2, at 3.9 mg/ml, which was considered a suitable concentration for crystallisation by the PCT results. Tiny cube-like microcrystals were observed after 1 month in one of the performed 96-well screens: MemGold HT-96. Two wells from the MemGold HT-96 screen had small crystals in, displayed in Figure 3.2.3a. For details on the crystallisation conditions, see Table 3.2.3a.

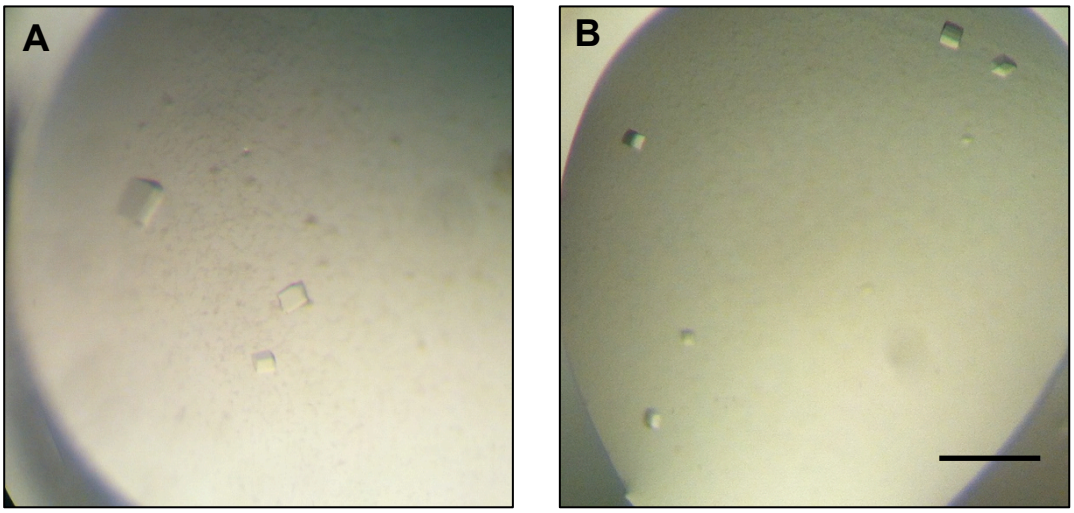


Figure 3.2.3a- Preliminary crystallisation hits of CDTb. A: MemGold HT-96 well B8, 3.9 mg/ml. B: MemGold HT-96 well H6, 3.9 mg/ml. Both wells contain crystals of CDTb purified in buffer 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 mM DTT. Scale bar represents 200 µm.

At least two crystals from each well of the initial hits were mounted in LithoLoops (Hampton Research) and taken to Diamond Light Source to test for signs of diffraction. For hits A and B from Figure 3.2.3a, mesh LithoLoops were used to ensure the crystals were scooped out of the wells successfully. One at a

time, the LithoLoops containing crystals were mounted on the goniometer and were rotated whilst being exposed to a beam of X-rays. No diffraction was noted for any of the mounted crystals from the wells pictured in Figure 3.2.3a.

If the crystals had been salt we would have observed the characteristic large diffraction spots that are usually noted during diffraction of salt crystals. The lack of diffraction spots observed indicates that the crystalline material is protein, rather than salt, but that the conditions required optimisation for diffraction studies.

Crystal	Screen and Well Details	Crystallisation Condition
A	MemGold HT-96 B8	0.1 M NaCl 0.02 M KMES pH 6.7 6.6 % PEG 4k
B	MemGold HT-96 H6	0.05 M lithium sulphate 0.1 M Tricine pH 7.4 7 % PEG 3k

Table 3.2.3a- Crystallisation conditions for preliminary crystallisation hits observed for CDTb. For crystal pictures see Figure 3.2.3a.

Optimisation of CDTb Crystals

As discussed in Chapter 2.1.1 under '*Optimisation of preliminary hits*`, the crystallisation reservoir commonly contains a buffer, salt and precipitant. The crystallisation conditions displayed in Table 3.2.3a are the corresponding conditions to the crystallisation hits displayed in the Figure 3.2.3a. In order to improve the crystal quality to a diffracting standard, the buffer, salt and precipitant concentrations were all varied. Initially for hit A in MemGold HT-96 B8, the precipitant PEG 4k concentration was varied between 0-10 % and the NaCl concentration was varied between 0-0.15 M. For hit B in MemGold HT-96 H6, the lithium sulphate concentration was varied between 0-0.1 M and the PEG 3k concentration was varied between 0-10 %. The stock solutions used for the optimisation trials were filtered using 0.4 µM pore size, 0.33 mm sterile filters (Millex) prior to use in crystallisation optimisations. The trials were performed on 24-well plates using both hanging and sitting drop methods, with a protein to reservoir ratio of 2:2 µl, at 16 °C. The optimisation plates were checked initially

after 24 hours and then fortnightly thereafter. After 1 month, crystals were observed in some of the wells on the optimisation plates. The crystals displayed in Figure 3.2.3b show a slightly different morphology to the preliminary hits. There was one flat square plate crystal and multiple non-cube-like crystals with rigid edges.

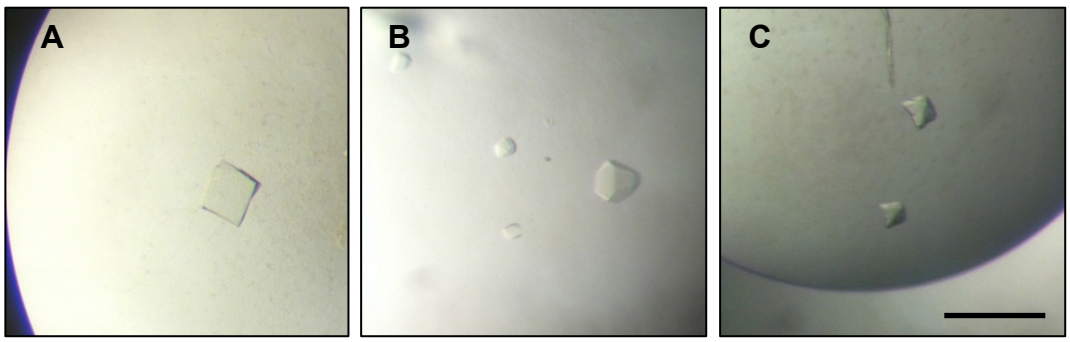


Figure 3.2.3b- Crystal hits of CDTb from optimisation trials, displaying different morphology to preliminary hits. Scale bar represents 200 µm.

The crystals were mounted in LithoLoops (Hampton Research) and taken to Diamond Light Source to test for signs of diffraction. As with the previous crystals no diffraction was noted for any of the mounted crystals from the optimisation trials. Due to the time restraints for the project, no further crystallisation trials were performed.

Crystallisation of CDTb'

Preliminary Screens

Preliminary screens for CDTb' were set up as discussed in Chapter 3.2.2, at 4.2 mg/ml, incubated at 16 °C, which was considered a suitable concentration for crystallisation by the PCT results. Tiny cube-like crystals were observed in one well of JCSG-plus HT-96, 2 months after the plate was set up, see Figure 3.2.3c-A. This was the only condition in which crystals grew during the preliminary screens (well E10 from JCSG-plus HT-96: 0.1 M Bicine pH 9.0 and 10 % PEG 6k). The crystals were mounted in mesh LithoLoops and taken to Diamond Light Source to test for signs of diffraction. One at a time, the

LithoLoops containing crystals were mounted on the goniometer and were rotated whilst being exposed to a beam of X-rays. No diffraction was noted for any of the mounted crystals from the well pictured in Figure 3.2.3c-A.

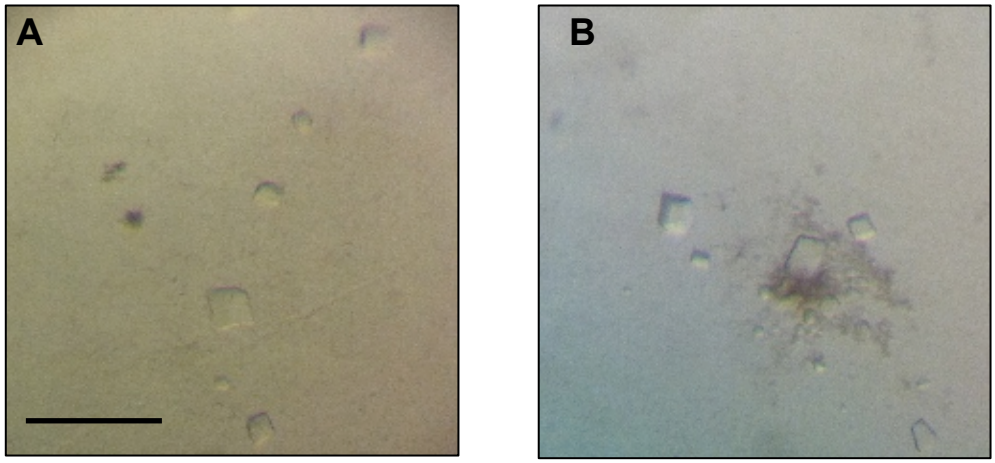


Figure 3.2.3c- Crystals of CDTb'. A: Preliminary crystallisation hits for CDTb' observed in one well on the JCSG-plus HT-96 screen. B: Crystallisation hits for CDTb' observed in the additives screen. Scale bar represents 200 µm.

Optimisation of CDTb' Crystals

As with the previous crystals for CDTb, attempts were made at optimising crystal quality in order to see diffraction. Initially, the concentration of PEG 6k was varied 0-30 %, followed by variation of the temperature at which the plates were incubated at: 4 °C, 16 °C and 21 °C. The concentration at which crystallisation plates were set up was also varied 2-15 mg/ml, in addition to the method of vapour diffusion used (hanging drop and sitting drop). An additives screen (Hampton Research) was also performed, whereby the initial crystallisation condition was used as the primary reservoir and then a concentration of 10 % of the additive reagents were added.

Crystals of CDTb' did not grow in any of the performed optimisations screens however, there was some success in crystal growth with the additive screen, where small cube-like crystals, of similar morphology to those grown previously, grew in the presence of 0.01 M betaine hydrochloride, see Figure 3.2.3c-B. These crystals were mounted in mesh LithoLoops and again were tested for signs of diffraction at Diamond Light Source. However, as with the original preliminary hits, no diffraction was observed.

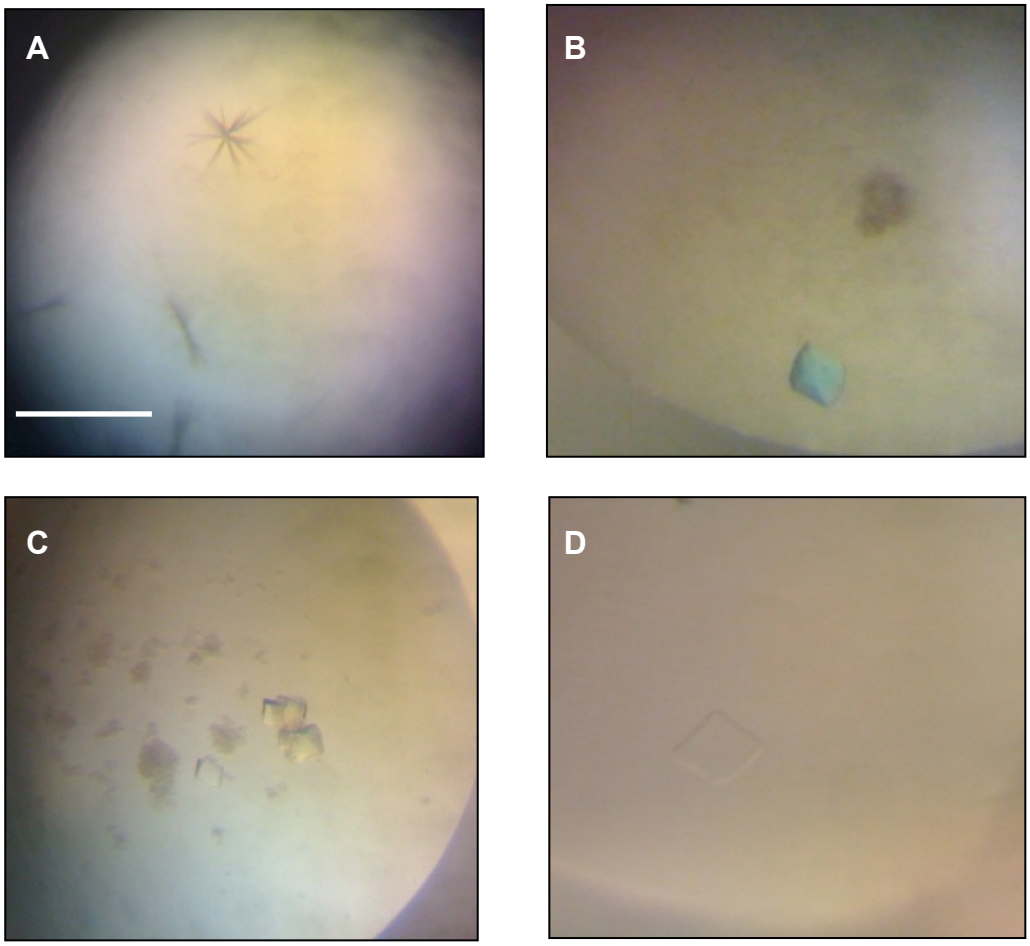


Figure 3.2.3d- CDTb' crystals from purification using buffer 50 mM Tris-HCl pH 7.5; 150 mM NaCl; 2 mM DTT and 0.2 % Tween-20. A: Small bundle like crystals observed in Structure screen I & II HT-96 screen, well B10. B: Unusual blue crystal observed in JCSG-plus HT-96 screen, well E10. C: Cluster of small cube-like crystals observed in JCSG-plus HT-96 screen, well C12. D: Thin plate-like crystal observed in Clear Strategy screen I HT-96 screen, well E11. Scale bar represents 200 µm.

As discussed in Chapter 3.1.3, the CDTb variant proteins all appear much larger than expected when purified using size-exclusion chromatography, which could be due to oligomerisation or aggregation. In case the proteins were aggregated, CDTb' was purified in the presence of 0.2 % Tween-20, in the usual buffer of 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 mM DTT, in attempts to prevent aggregation. The purified protein was concentrated to 4 mg/ml and was screened as usual for crystal growth. After 3 months, crystals were observed in the Structure screen I & II HT-96 screen, the JCSG-plus HT-96 screen and the

Clear Strategy I HT-96 screen, see Figure 3.2.3d for crystal pictures and Table 3.2.3b for details of crystallisation conditions.

Crystal	Screen	Crystallisation Conditions	Optimisation Range
A	Structure Screen I & II HT-96 well B10	0.2 M magnesium chloride 0.1 M sodium HEPES pH 7.5 30 % 2-propanol	0-0.3 M magnesium chloride
			0.1 M sodium HEPES pH 7.5 kept constant
			0-40 % 2-propanol
B	JCSG-plus HT-96 well C12	0.1 M Bicine pH 9.0 10 % PEG 6k	0.1 M Bicine pH 9.0 kept constant
			2.5-15 % PEG 6k
C	JCSG-plus HT-96 well E10	10 % PEG 1k 10 % PEG 8k	5-15 % PEG 1k
			5-15 % PEG 8k
D	Clear Strategy screen I HT-96 well E11	0.2 M potassium thiocyanate 0.1 M Tris pH 7.5 15 % PEG 4k	0.2 M potassium thiocyanate kept constant
			0.1 M Tris pH 7.5 kept constant
			5-30 % PEG 4k

Table 3.2.3b- the details of the crystallisation conditions for crystal hits displayed in Figure 3.2.3d, along with the range of concentrations used during optimisations trials.

The small bundles of crystals observed in Figure 3.2.3d-A were too fine to collect data from; therefore they were used for seeding, which will be explained later. The small blue crystal in Figure 3.2.3d-B, was observed in the same condition as the original CDTb' crystal hit from the preliminary screens from JSCG-plus HT-96 E10. The crystal was stuck to the bottom of the well, so 1 µl of reservoir solution was added to the drop to aid in crystal mounting. However, the crystal dissolved completely upon addition of reservoir solution. The crystals in Figure 3.2.3d-C were mounted in LithoLoops and taken to Diamond Light Source but showed no signs of diffraction and finally the crystal in Figure 3.2.3d-D dissolved when attempting to mount in a LithoLoop.

Table 3.2.3b describes the variations in crystallisation conditions used during the optimisation attempts. The buffer in each of the conditions was kept constant whilst the PEG and salt concentrations were varied. After six months there was still no sign of crystal growth.

Microseeding plates were set up as discussed in Chapter 2.1.1, using the bundle of fine crystals displayed in Figure 3.2.3d-A, in attempts to improve crystal morphology. In brief, the crystals in the well were crushed until no longer visible using a needle. The drop was diluted in 2 µl of reservoir and was aspirated and dispensed into a small Eppendorf tube. The tube was mixed by vortex for 5 min at room temperature to ensure complete crushing of the crystals. A 24-well plate was set up using the same batch of protein that produced the small crystal bundles in Figure 3.2.3d-A, with the original crystallisation condition. The hanging drop method was used with a drop size ratio of 1 µl: 1 µl. The plate was incubated at 16 °C for 3 hours prior to microseeding. The solution containing crushed crystal fragments was then diluted in series as displayed in Table 3.2.3c, which represents a 24-well plate and 1 µl of the diluted seeds was added to each of the pre-equilibrated drops. The microseeded plate was then incubated at 16 °C and checked regularly for crystal growth.

	1	2	3	4	5	6
A	1:50	1:50	1:50	1:100	1:100	1:100
B	1:200	1:200	1:200	1:400	1:400	1:400
C	1:500	1:500	1:500	1:5000	1:5000	1:5000
D	1:10000	1:10000	1:10000	1:25000	1:25000	1:25000

Table 3.2.3c- Dilution series for 24-well microseeding plate. Small bundles of crystals were crushed and diluted as in the table. The diluted CDTb' crystals were added to drops pre-equilibrated with 1 µl: 1 µl ratio of protein to reservoir, to make the final ratio 2 µl: 1 µl of protein to reservoir.

After 3 months, more bundles of fine crystals were observed on the microseeded optimisation plate, which were unsuitable for data collection, see Figure 3.2.3e-A. Therefore as used previously, the Additive Screen (Hampton Research) was performed using the successful crystallisation conditions for the crystals in Figure 3.2.3d-A, B, C and D as the primary reservoir. Additionally the Silver Bullet screen was performed in the same manner. There was one successful crystallisation hit which was observed on the Additive Screen plate that was set up using the crystallisation condition form JCSG-plus HT-96 well

E10 which had shown small blue crystal growth previously. The crystals are displayed in Figure 3.2.3e-B and were grown in the presence of the additive 0.01 M betaine hydrochloride. During attempts to mount the crystals in LithoLoops, the crystals were fixed to the bottom of the well, so an additional 2 µl of reservoir solution was added to help with mounting. Unfortunately the crystals dissolved upon addition of the reservoir. This indicated that the crystals were likely to be protein, and so optimisation plates were set up where the original reservoir of 10 % PEG 1k and 10 % PEG 8k was kept constant and a range of 0-0.1 M betaine hydrochloride was trialled. As yet, no crystals have been observed in the betaine hydrochloride optimisation trial.

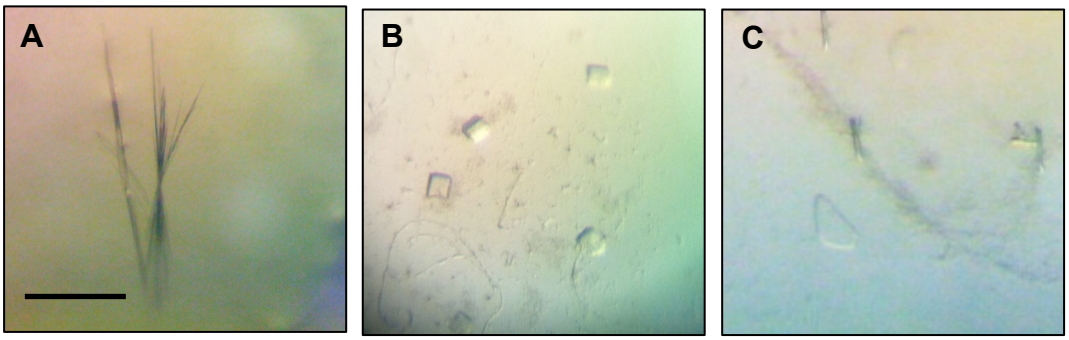


Figure 3.2.3e- Crystals of CDTb' grown using various optimisation methods. A: Bundles of fine crystals grown from the microseeding plate, showing no improvement of crystal morphology. B: Small cube-like crystals observed from the Additive Screen, grown in the presence of betaine hydrochloride. C: Alternative triangular morphology crystals observed 6 months after preliminary screens were performed using 0.2 % Tween-20 in the purification buffer. Scale bar represents 200 µm.

As discussed earlier, the preliminary screens were performed using CDTb' purified in the standard buffer with the addition of 0.2 % Tween-20. After 3 months, crystals were observed as displayed in Figure 3.2.3d, however after 6 months, one other crystallisation hit was observed in the Clear Strategy Screen I HT-96 well F4, displaying in Figure 3.2.3e-C. These crystals had a triangular morphology, which is different to all of the previously observed crystal hits. The crystal was mounted in a LithoLoop and taken to Diamond Light Source to test for signs of diffraction. The loop showed a large quantity of ice formation and the crystal was not visible in the loop. Therefore, attempts were made at reproducing

more crystals in this specific condition so that a broad range of cryo-protectants could be trialled to reduce the ice problem, however there have been no indications of crystal growth as yet.

In addition, optimisation plates were set up using the same technique as previously. The original crystallisation condition contained 0.2 M potassium bromide, 0.1 M Tris pH 7.5, 10 % PEG 8k, 10 % PEG 1k. Therefore, Tris buffer was kept constant, potassium bromide was varied 0-0.5 M, PEG 8k was varied 0-30 % and PEG 1k was varied 0-30 %. These optimisations were performed towards the end of the project, and until date there have been no crystals observed for this batch of optimisations, however the preliminary crystals took 6 months to grow, and so this time frame is yet to be reached.

Concurrently, an alternative method to produce diffraction quality crystals was used whereby the preliminary 96-well screens were performed with the protein in different buffer systems. This method is used to extend the range of pH buffers the protein is subjected to during crystallisation. In the final stages of purification of CDTb', the pure protein was dialysed separately into the following three buffers, using 50 x volume at 4 °C:

1. Buffer 1: 50 mM MES pH 6.5; 150 mM NaCl; 2 mM DTT
2. Buffer 2: 50 mM HEPES pH 7.0; 150 mM NaCl; 2 mM DTT
3. Buffer 3: 50 mM Bicine pH 8.5; 150 mM NaCl; 2 mM DTT

After dialysis, CDTb' was concentrated to 4 mg/ml and the crystallisation screens JCSG-plus HT-96, Clear Strategy Screen I HT-96, Clear Strategy Screen II HT-96, MIDAS HT-96, and Structure screen I + II HT-96 (Molecular Dimensions) were performed and incubated at 16 °C. Samples of the CDTb' purification, and of the buffer exchanges were analysed using 12 % Tris-glycine SDS-PAGE, see Figure 3.2.3f. By setting up crystallisation screens using a variety of protein buffers, there should be a greater variation in crystallisation conditions during the screening. The gel picture in Figure 3.2.3f, shows the buffer exchanged protein in lanes 7, 8 and 9. The protein looks to be stable in the three alternative buffers, although there appears to be more smaller bands below the CDTb' band, this could be a concentration effect rather than unstable protein.

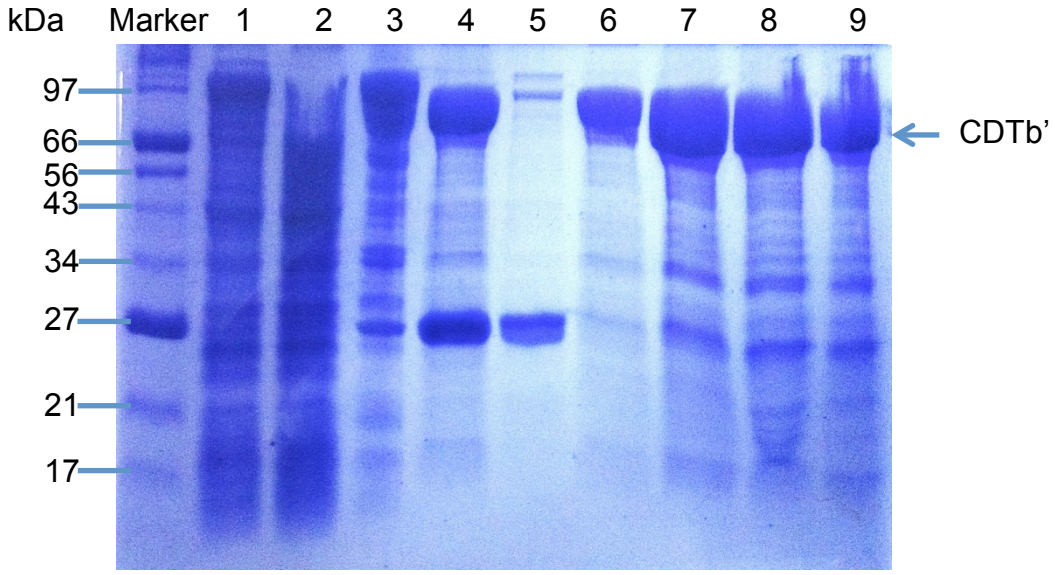


Figure 3.2.3f- 12 % Tris-glycine SDS-PAGE gel picture displaying the purification of CDTb' and the subsequent buffer exchanges. **Lane 1: Loaded cell lysate onto GSTrap column. Lane 2: Non-bound protein flow-through. Lane 3: Bound CDTb'GST eluted from the GSTrap column using 20 mM reduced glutathione. Lane 4: Cleavage of GST-tag using PreScission Protease. Lane 5: GST-tag removed. Lane 6: Final pure CDTb' in buffer pH 8.0. Lane 7: Dialysis into pH 6.5 buffer 1. Lane 8: Dialysis into pH 7.0 buffer 2. Lane 9: Dialysis into pH 8.5 buffer 3.**

Two months after the new screens were performed with the alternative buffers, two crystal hits were observed. There were micro-crystals noted for protein in buffer 3 at pH 8.5, see Figure 3.2.3g-A, and one larger crystal had grown for the protein buffer 1 at pH 6.5, see Figure 3.2.3g-B. Crystals from Figure 3.2.3g-A were mounted in mesh LithoLoops and were taken to Diamond Light Source to test for diffraction. As expected for crystals of this size, following a similar pattern to previous crystals, no signs of diffraction were observed. The crystal in Figure 3.2.3g-B was noted in the Clear Strategy Screen I HT-96 screen, well F3, containing the following crystallisation condition: 0.1 M Tris pH 7.5, 0.2 M magnesium chloride, 10 % PEG 8k and 10 % PEG 1k. The crystal was mounted in a LithoLoop and taken to Diamond Light Source to test for signs of diffraction. The LithoLoop containing the crystal was mounted on a goniometer and was rotated whilst being exposed to a beam of X-rays. The crystal showed some small spots at very low resolution (approximately 50 Å) but after exposure to the beam for a short period of time the crystal lost its diffracting ability. This is the first concrete evidence obtained from these trials that these crystals are protein and not salt.

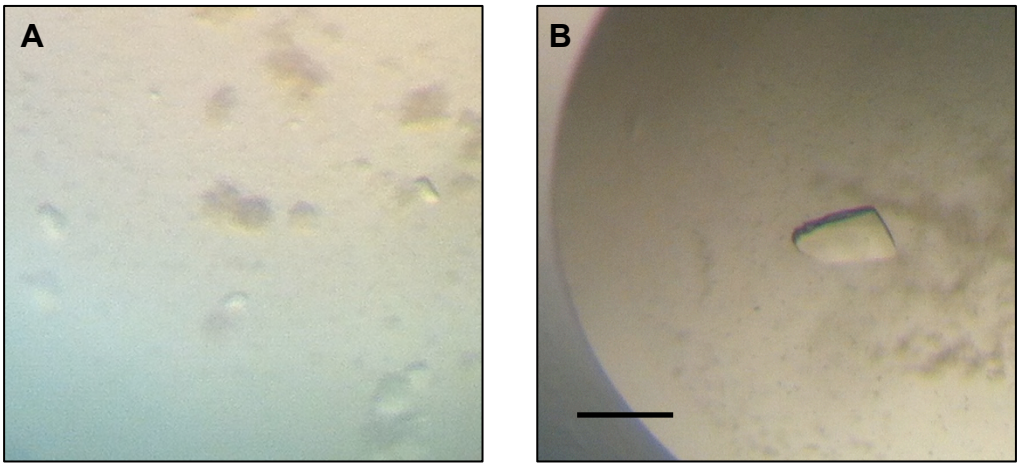


Figure 3.2.3g- Crystal hits of CDTb' from the optimisation trials after the buffer exchanges. A: Crystals hits for CDTb' in buffer 3 pH 8.5. B: Large crystal hit for CDTb' in buffer 1 pH 6.5. Scale bar represents 200 µm.

In order to improve the quality of the CDTb' crystal, optimisation trials were performed whereby the individual components of the successful crystallisation condition were varied. The Tris buffer was kept consistent whilst the magnesium chloride concentration was varied 0-25 M, the PEG 8k was varied 0-25 % and the PEG 1k was also varied 0-25 %. Both sitting and hanging drop methods of vapour diffusion were used in addition to varying the protein concentration 2-10 mg/ml.

In addition to the above screens and optimisation trials, the Additive screen and Silver bullet screen were performed, using the same batch of protein that the diffracting crystal was grown from (buffer 1 pH 6.5), and the Clear Strategy Screen I F3 reservoir as purchased from Molecular Dimensions. These additional screens were performed at 2 mg/ml, 4 mg/ml, 6 mg/ml and 8 mg/ml, using the sitting drop method and were incubated at 4 °C, 16 °C and 21 °C, to ensure maximum chance of crystal growth.

After approximately one week, crystals were observed on the Additive screen optimisation plate in positions A5 and E3 which contained the additives copper (II) chloride dihydrate and dextran sulphate sodium salt respectively. The morphologies of these crystals varied significantly from the original large crystal, see Figure 3.2.3h-A-D.

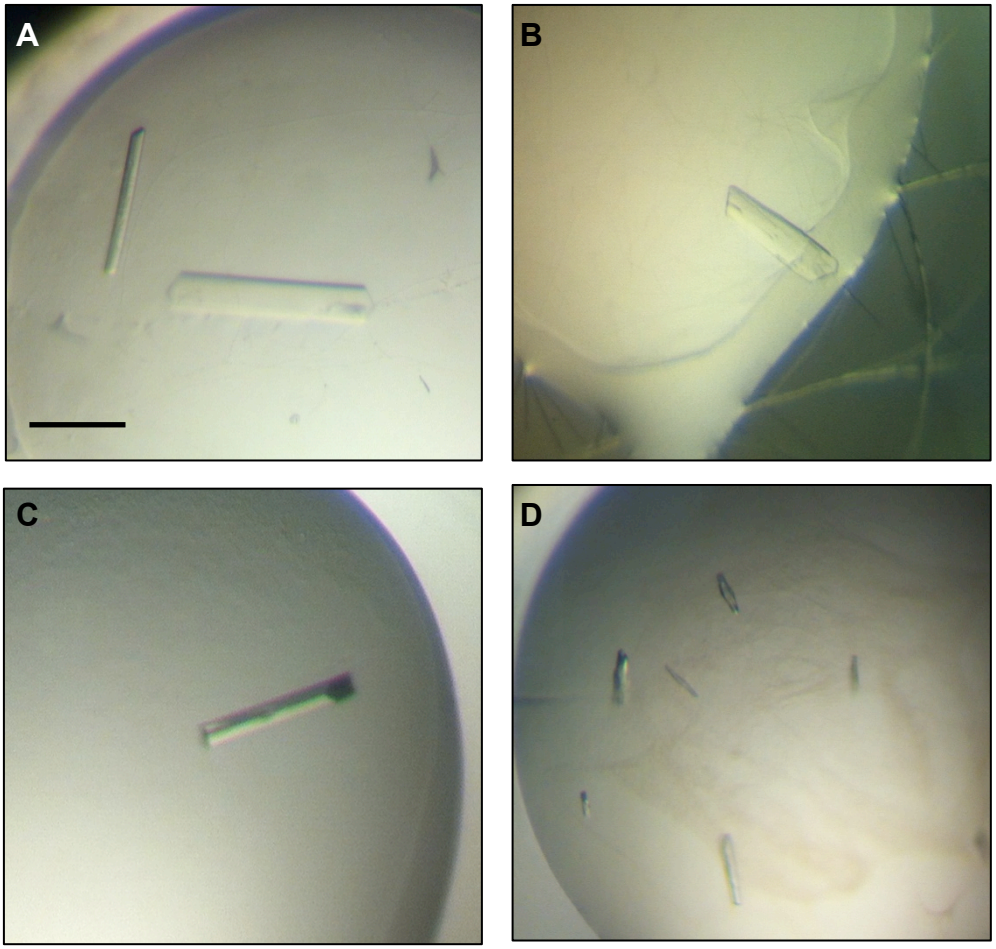


Figure 3.2.3h- Crystals of CDTb' observed during additive screen optimisation trials. A: Crystals were grown at 4 mg/ml in dextran sulphate sodium salt. B: Crystal was grown at 6 mg/ml in dextran sulphate sodium salt. C: Crystal was grown at 2 mg/ml in dextran sulphate sodium salt. D: Crystals were grown at 2 mg/ml in copper (II) chloride dihydrate. All the above crystals were grown at 16 °C. Scale bar represents 200 µm.

All of the crystals were mounted in LithoLoops (Hampton Research) and taken to Diamond Light Source to test for signs of diffraction. Crystal hits A, B and C from Figure 3.2.3h disappointingly turned out to be salt crystals, most likely of dextran sulphate sodium salt. Crystal D however diffracted to 8 Å, which is a significant improvement on the last crystal that diffracted. However, the crystal suffered radiation damage from the beam before enough data could be collected for processing. Figure 3.2.3i displays the image of the crystal mounted in the LithoLoop (A), with the beam position marked as a red circle with a cross in the middle, and a diffraction image (B) with a close up (C).

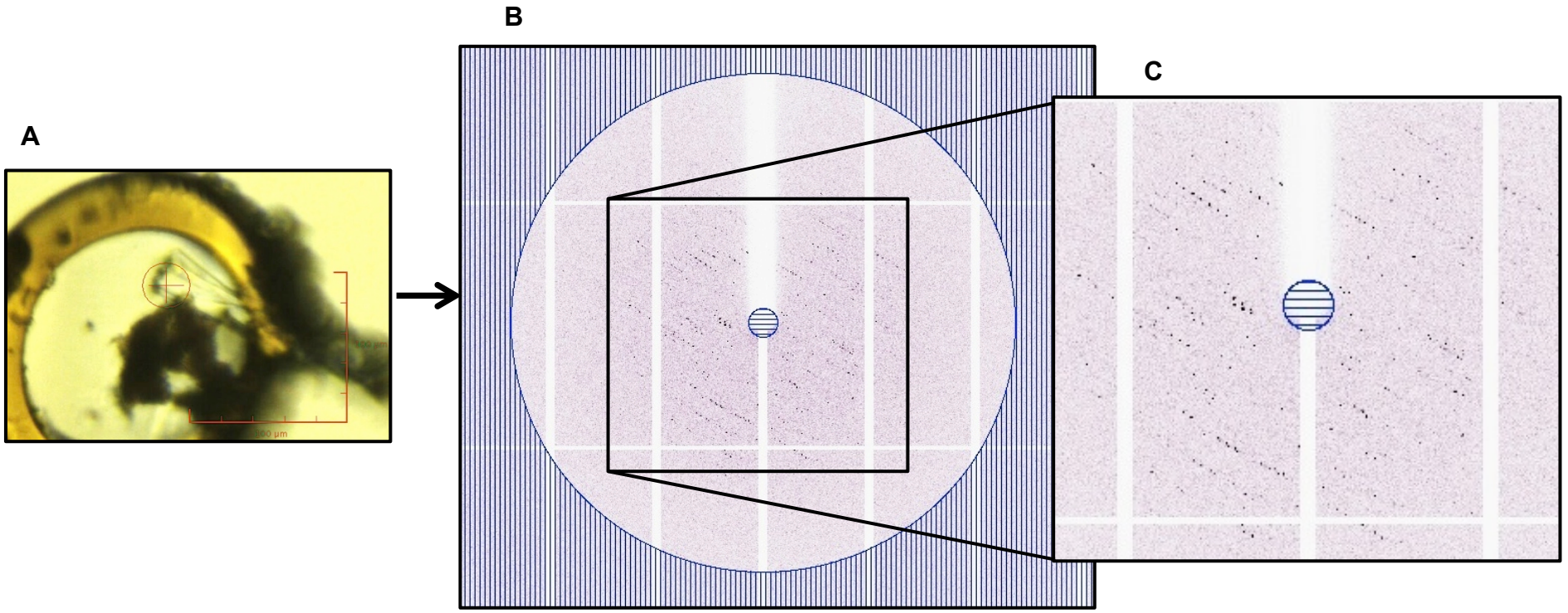


Figure 3.2.3i- CDTb' diffraction. A: Crystal of CDTb' mounted in LithoLoop during exposure to beam at Diamond Light Source with red circle indicating beam position. B: A diffraction image collected for CDTb' displaying the 5 Å resolution ring. C: A closer look at the diffraction spots of CDTb' indicating a long cell dimension.

CDTb' Data Analysis

Data were collected at wavelength (λ) 0.9795 Å, 50 % transmission, with distance from the detector set to 1 m. Data were collected at 0 °, 45 ° and 90 ° oscillation angles, with the crystal subjected to 10 s of beam exposure at each angle. Diffraction spots observed for one of the images are displayed in Figure 3.2.3i-B and C. Attempts were made at indexing the data using Mosflm, in addition to the autoprocessing tools at Diamond Light Source; xia2 and fast_dp (Winter, 2010). However, the data could not be indexed and so cell dimensions could not be obtained.

Crystallisation of CDTb''

Preliminary Screens

Purified CDTb'' in buffer 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT and 10 % glycerol was used to set up crystallisation trials. Preliminary screens for CDTb'' were set up as discussed in Chapter 3.2.2, at 5 mg/ml, which was considered a suitable concentration for crystallisation from the PCT result. The plates were checked after 24 hours and weekly thereafter. After approximately 2 months, tiny cube-like crystals were observed in the JCSG-plus HT-96 wells E1 and F10, and ProPlex HT-96 well H4, see Figure 3.2.3j.

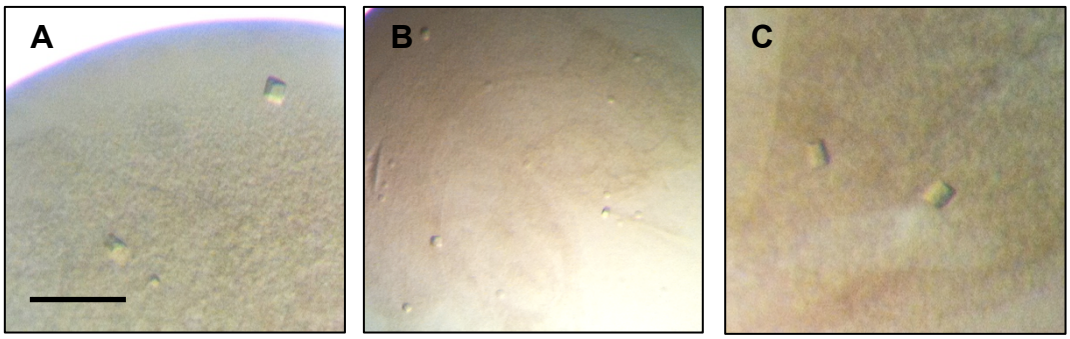


Figure 3.2.3j- CDTb'' preliminary crystallisation hits. A: Small cube-like crystals observed in JCSG-plus HT-96 well E1. B: Tiny cube-like crystals observed in JCSG-plus HT-96 well F10. C: Slightly larger cube-like crystals observed in ProPlex HT-96 well H4. Scale bar represents 200 µm.

The small cube-like crystals were mounted in mesh LithoLoops and taken to Diamond Light Source for signs of diffraction. Unfortunately none of the crystals mounted showed signs of diffraction and so optimisation trials were performed.

Optimisation of CDTb'' Crystals

24-well crystallisation plates were set up using both the sitting and hanging drop methods. The plates were incubated at 16 °C and monitored after 24 hours in the first instance and then weekly thereafter. The variations in crystallisation condition content tested are displayed in Table 3.2.3d. After 2 months, crystals were observed in one well of the optimisation plate, which happened to be the original, non-varied condition from JCSG-plus HT-96 well F10.

Crystal	Screen	Crystallisation Condition	Optimisation Trials
A	JCSG-plus HT-96 well E1	0.1 M sodium cacodylate pH 6.5	0.1 M sodium cacodylate pH 6.5 kept constant
		1.0 M tri-sodium citrate	0-2.0 M tri-sodium citrate
B	JCSG-plus HT-96 well F10	1.1 M sodium malonate	0-2.0 M sodium malonate
		0.1 M HEPES pH 7.0	0.1 M HEPES pH 7.0 kept constant
		0.5 % Jeffamine ED-2001	0-1 % Jeffamine ED-2001
C	ProPlex HT-96 well H4	1.4 M sodium malonate	0-2.0 M sodium malonate Alternative of 0-2.0 M sodium formate

Table 3.2.3d– Crystallisation hit conditions from crystals in Figure 3.2.3j and the variations in concentration tested during optimisation attempts.

On checking the preliminary screens on a weekly basis, more crystals were observed on the ProPlex HT-96 screen in wells F12, G9, G10, and H5, 3 months after the plates were initially set up, see Figure 3.2.3k. The small cube-like crystals were mounted in mesh LithoLoops and taken to Diamond Light Source for signs of diffraction. Unfortunately none of the crystals mounted showed signs of diffraction, therefore optimisation trials were performed using 24-

well plates, with both the sitting and hanging drop methods, using the standard protocol discussed previously of varying salt and precipitant concentrations initially, see Table 3.2.3e for details.

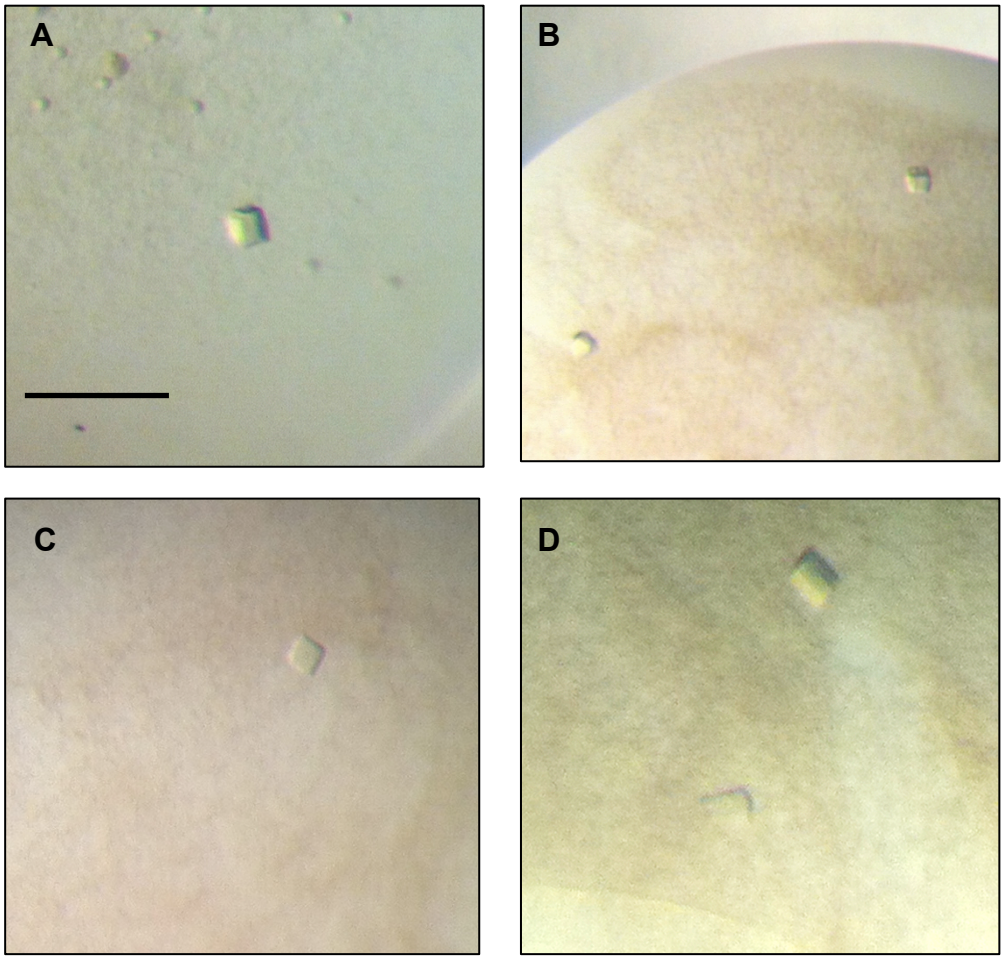


Figure 3.2.3k- CDTb'' crystals all grown in various wells on the ProPlex HT-96 crystallisation screen. A: Small cube-like crystal grown in ProPlex HT-96 well F12. B: Smaller cube-like crystals grown in ProPlex HT-96 well G9. C: Small cube-like crystal grown in ProPlex HT-96 well G10. D: Two slightly larger cube-like crystals grown in ProPlex HT-96 well H5. Scale bar represents 200 µm.

After three months there was no sign of crystal growth in any of the optimised conditions. Therefore, CDTb'' was purified en masse over the period of two months to gain enough protein to set up a range of different screens. In addition, a range of salt concentrations of 50 mM, 100 mM and 150 mM were used in the purification buffer.

All of the preliminary screens discussed in Chapter 3.2.1 were performed at a large range of concentrations: 2 mg/ml, 3 mg/ml, 4 mg/ml, 5 m/ml, 6 mg/ml, 7

mg/ml, 8 mg/ml, 9 mg/ml and 10 mg/ml, at three different temperatures: 4 °C, 16 °C and 21 °C. After three months, there were some signs of crystal growth observed on the Structure screen I & II HT-96, which was set up at 6 mg/ml and incubated at 16 °C, see Figure 3.2.3l. The crystals shown in Figure 3.2.3l were mounted in mesh LithoLoops and taken to Diamond Light Source to test for signs of diffraction. No diffraction was observed for any of the crystals.

Crystals	Screen and Well	Crystallisation Condition	Optimisation Trials
A	ProPlex HT-96 well F12	0.1 M sodium HEPES pH 7.0	0.1 M sodium HEPES pH 7.0 kept constant
		1.5 M ammonium sulphate	0-2.0 M ammonium sulphate
B	ProPlex HT-96 well G9	1.3 M sodium potassium hydrogen phosphate pH 7.0	0-2.0 M sodium potassium hydrogen phosphate pH 7.0 kept constant
C	ProPlex HT-96 well G10	1.6 M sodium potassium hydrogen phosphate pH 6.5	0-2.0 M sodium potassium hydrogen phosphate pH 6.5 kept constant
D	ProPlex HT-96 well H5	0.1 M Tris pH 8.0	0.1 M Tris pH 8.0 kept constant
		1.2 M sodium potassium tartrate	0-2.0 M sodium potassium tartrate

Table 3.2.3e– Crystallisation hit conditions for crystals from Figure 3.2.3k and the variations in concentration trialled during optimisation attempts.

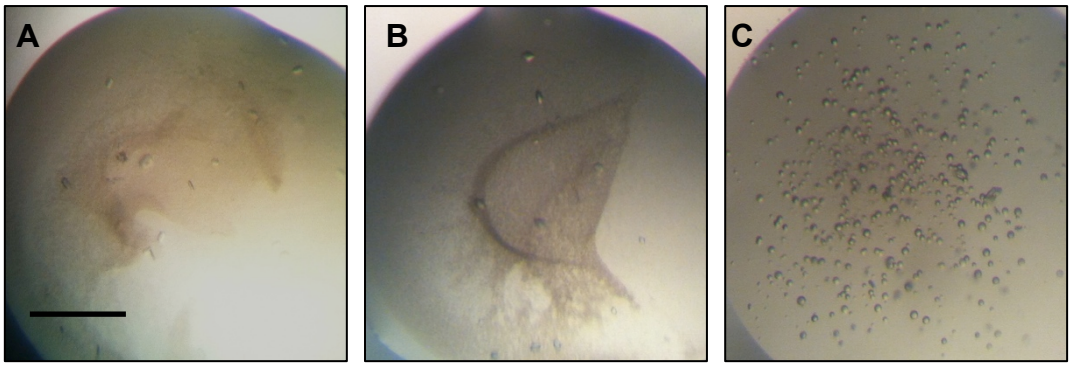


Figure 3.2.3l- Preliminary signs of CDTb'' crystallisation. A: Crystals grown in 0.2 M magnesium chloride, 0.1 M sodium HEPES pH 7.5 and 30 % PEG 400. B: Crystals grown in 0.2 M magnesium chloride, 0.1 M Tris pH 8.5 and 30 % PEG 4k. C: Crystals grown in 0.2 M magnesium formate. Scale bar represents 200 µm.

As the aforementioned screens and optimisations were unsuccessful up until this point, the next steps taken to produce diffraction quality crystals was to repeat the preliminary 96-well screens using protein in different buffer systems.

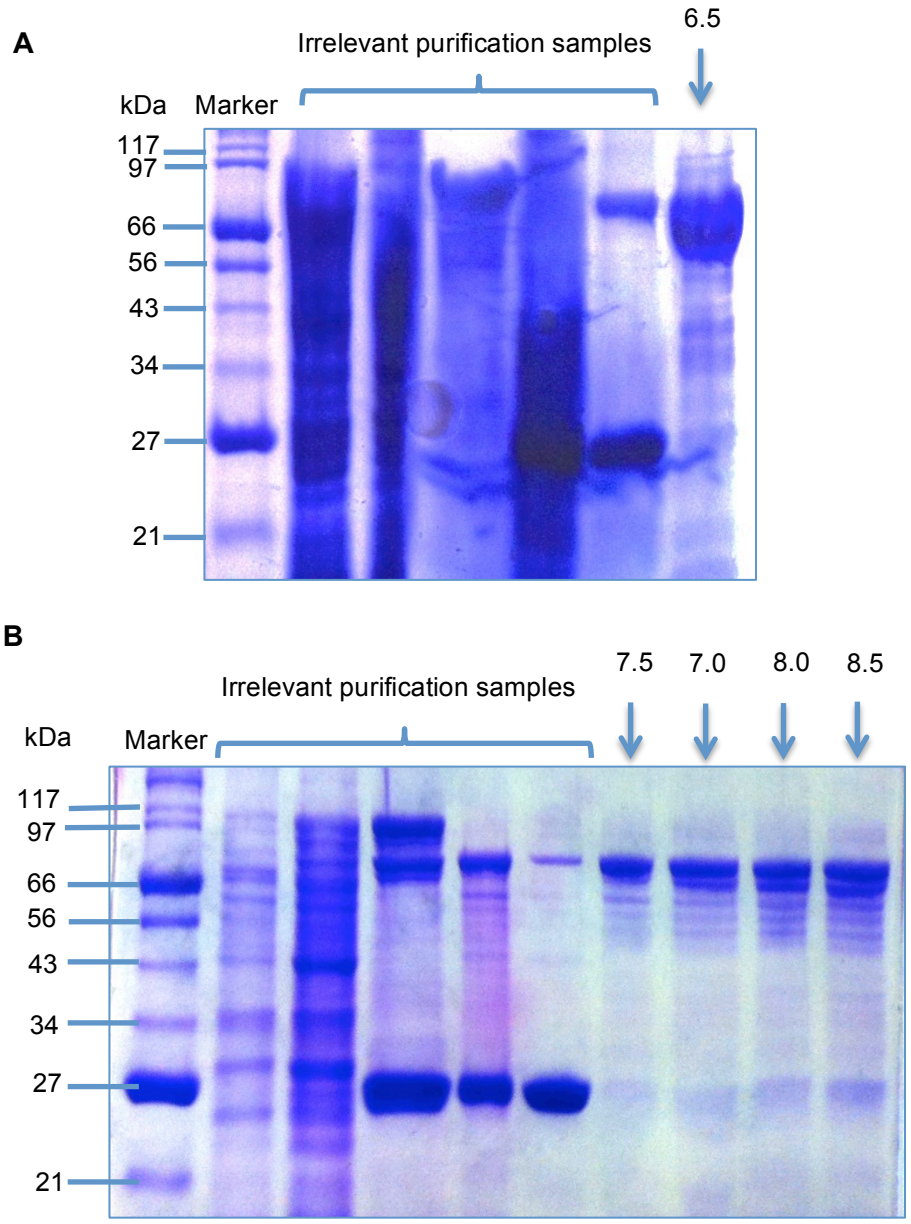


Figure 3.2.3m- A 12% Tris-glycine SDS-PAGE gel pictures displaying pure CDTb'' in the various buffers. A: CDTb'' purification using standard GST-affinity chromatography with final protein in Buffer 1, pH 6.5. B: CDTb'' purification using standard GST-affinity chromatography at pH 7.5 with final pure protein dialysed into Buffers 2 (pH 7.0), 3 (pH 8.0) and 4 (pH 8.5).

In the final stages of purification of CDTb'', the pure protein was dialysed separately into the following four buffers, using 50 x volume buffer at 4 °C:

1. Buffer 1: 50 mM MES pH 6.5; 150 mM NaCl and 2 mM DTT
2. Buffer 2: 50 mM HEPES pH 7.0; 150 mM NaCl and 2 mM DTT
3. Buffer 3: 50 mM Tris pH 8.0; 150 mM NaCl and 2 mM DTT
4. Buffer 4: 50 mM Bicine pH 8.5; 150 mM NaCl and 2 mM DTT

The purity of CDTb'' in the various buffers was analysed by 12 % Tris-glycine SDS-PAGE, see Figure 3.2.3m. After CDTb'' was deemed suitably pure in the range of buffers, the protein was concentrated to 5 mg/ml and some of the preliminary screens were performed at 16 °C: Clear Strategy Screen I HT-96, Clear Strategy Screen II HT-96, JCSG-plus HT-96, Morpheus HT-96, MIDAS HT-96 and Structure Screen I & II HT-96. This technique was used because the range of buffers will slightly alter the buffer in the crystallisation condition during screening, which should give a wider variety of alternate conditions during screening. This technique was also used with no buffer whereby CDTb'' was dialysed into HPLC-grade water and 150mM NaCl, the protein was concentrated to 2 mg/ml and all of the preliminary screens available were performed and incubated at 16 °C.

After two months, one large piece of crystalline matter was noted in one well from the JCSG-plus HT-96 screen, well G6, and small fragments were noted in the screen JCSG-plus HT-96, well B12. Both wells contained crystals from protein purified in HPLC-grade water and 150 mM NaCl, at a concentration of only 2 mg/ml, see Figure 3.2.3n.

Crystals displayed in Figure 3.2.3n-B were fused to the bottom of the well and so could not be mounted. The crystals dissolve during attempts to mount in LithoLoops. The large crystal from Figure 3.2.3n-A was mounted in a LithoLoop and taken to Diamond Light Source to test for signs of diffraction. For the first time, diffraction spots were observed up to 8 Å for the CDTb'' crystal, however upon exposing the crystal to the beam for more than a few seconds, the crystal suffered radiation damage, losing its diffraction quality

.

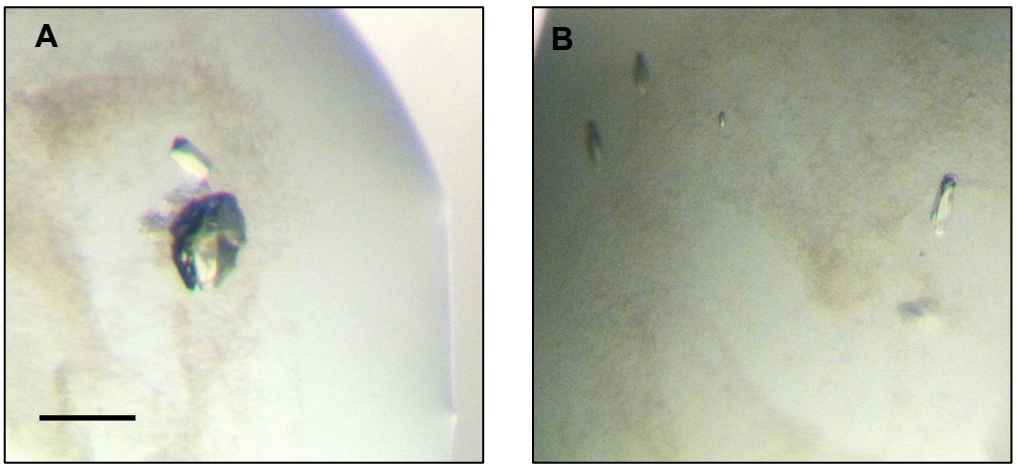


Figure 3.2.3n- Crystals of CDTb'' from protein purified using HPLC-grade water and 150 mM NaCl and no buffer. A: Crystals of CDTb'' from screen JCSG-plus HT-96, well G6. B: Crystals of CDTB'' from screen JCSG-plus HT-96, well B12. Scale bar represents 200 µm.

Table 3.2.3f displays the conditions in which CDTb'' crystals grew, and the variations in the components tested for optimisation trials. The original crystallisation conditions were also used as the primary condition to which the Additive screen and Silver Bullet screen were added. These primary conditions were purchased from Molecular Dimensions to ensure no variations by in-house production. The purchased buffers were used to set up three 96-well plates of identical conditions, for each of the three different temperature incubators, 4 °C, 16 °C and 21 °C, in attempts to reproduce the original diffracting crystals.

Crystal	Screen and Well	Crystallisation Condition	Conditions Trialled
A	JCSG-plus HT-96 well G6	0.2 M sodium malonate pH 7.0	0-0.3 M sodium malonate pH 7.0
		20 % PEG 3350	0-25 % PEG 3350
B	JCSG-plus HT-96 well B12	0.2 M tri-potassium citrate	0-0.3 M tri-potassium citrate
		20 % PEG 3350	0-25 % PEG 3350

Table 3.2.3f– Crystallisation hit conditions for crystals from Figure 3.2.3i and the variations in concentration tested during optimisation trials.

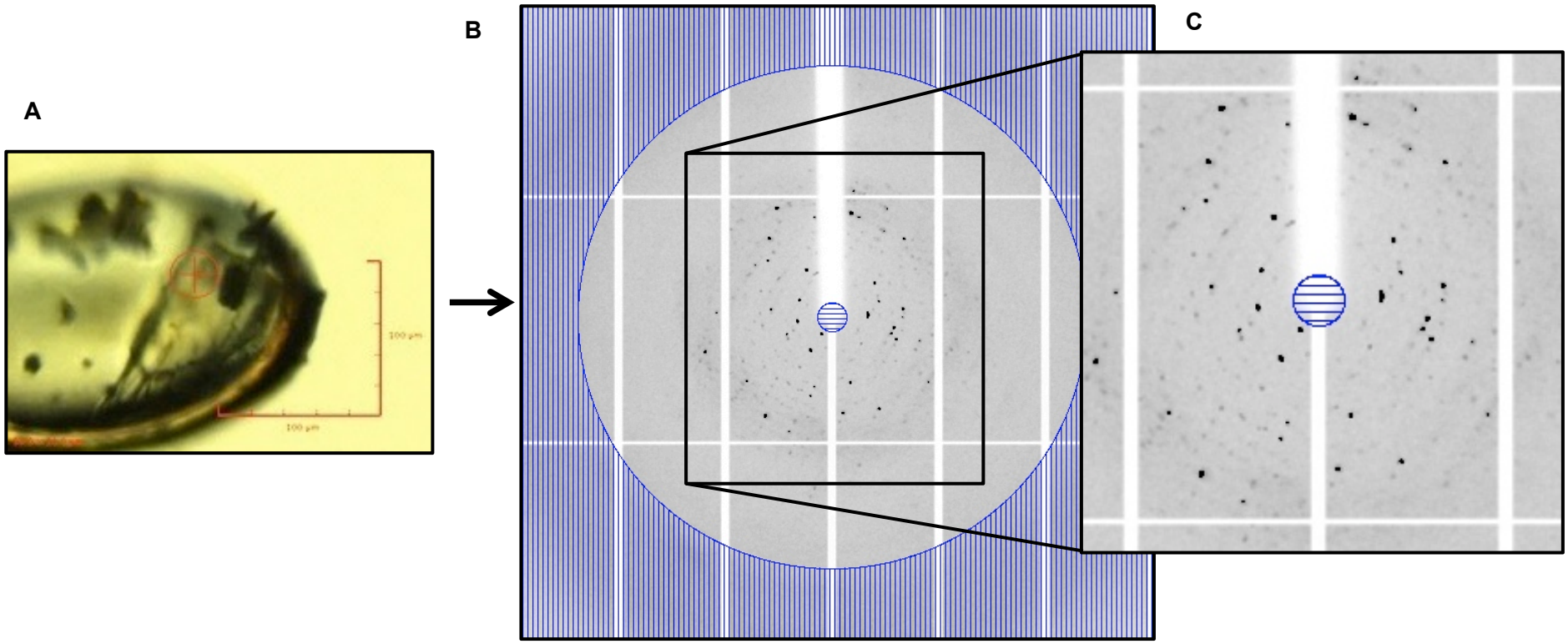


Figure 3.2.3o- CDTb'' diffraction. A: Crystal of CDTb'' mounted in LithoLoop during exposure to beam at Diamond Light Source with red circle indicating beam position. B: A diffraction image collected for CDTb'' displaying the 5 Å resolution ring . C: A closer look at the diffraction spots of CDTb'' showing some mosaicity within the crystal.

In addition, 24-well and 96-well plates were set up using these purchased buffers at a protein concentration range of 2 mg/ml, 4 mg/ml, 6 mg/ml, 8 mg/ml and 10 mg/ml and a temperature range of 4 °C, 16 °C and 21 °C. The 96-well plates were set up using the sitting drop method whereas the 24-well plates were set up using both sitting and hanging drop methods.

These optimisations were performed towards the end of the project, and as yet crystals have not appeared in the wells, but due to the extensive range of crystallisation environments set up, the likelihood of CDTb'' crystal formation is amplified.

Analysis of the First Diffraction Spots Observed for CDTb''

As with the CDTb' crystal, data were collected at wavelength (λ) 0.9686 Å, 20 % transmission on the I24 beam line at Diamond Light Source, with distance from the detector set to 1 m. At each oscillation of 1 ˚, the crystal was exposed for 2 seconds and 3 images were collected. Diffraction spots observed for one of the images are displayed in Figure 3.2.3o-B and C. The crystal stopped diffracting shortly after the third image was taken.

The data were indexed using Mosflm, to give the following cell dimensions: a= 173 Å b= 173 Å c= 126 Å and α= 90 ˚ β= 90 ˚ γ= 120 ˚, the space group prediction is P3. The Matthew's coefficient and solvent content were calculated to be 2.72 and 54.8 % respectively with a predicted 6 molecules in the asymmetric unit.

3.3 Low-Resolution Structural Analysis of the CDTb Variants

3.3.1 Introduction

CDTb

As discussed in previous chapters, *C. difficile* produces an ADP-ribosylating toxin, amongst other potent exotoxins. The ADP-ribosylating toxin, CDT is made up of two components, CDTa (enzymatic component) and CDTb (transport component). The main focus of Chapter 3 was to characterise CDTb variants at a structural level in order to understand the preparations that CDTb undertakes prior to uptake into the cells. The CDTa component has been characterised structurally, however the CDTb component has proven difficult to work with in terms of crystallography techniques, see Chapter 3.2. As discussed previously, each of the forms of CDTb have been crystallised, two of which have shown diffraction (CDTb' and CDTb''). However, the quality of the crystals is poor and in order to solve the three-dimensional structure at high resolution, better quality crystals are required. Extensive attempts at improving crystal quality have not as yet been successful. Therefore in this section a number of low-resolution structural analysis techniques have been employed to decipher some three-dimensional aspects of CDTb variants.

Dynamic Light Scattering

Dynamic Light Scattering (DLS) is a technique used to determine the size of proteins, nucleic acids, and complexes, and to detect aggregates in macromolecular solutions. However, DLS will be used in this Chapter to determine whether CDTb variants are monodisperse or polydisperse, hence whether or not the CDTb samples are suitable for SAXS data collection. For details on the theory of DLS see Chapter 2.2.1.

Small-Angle X-ray Scattering

Small-angle X-ray scattering (SAXS) is a technique used to determine low-resolution structures of macromolecules in solution, which can provide information about the size, distribution and overall shape of the macromolecule. The advantage of using SAXS over crystallography is that crystalline matter is not required, which is often a limiting step in X-ray crystallography. A small sample of pure monodisperse protein is all that is required for SAXS, however there are limitations, the key restriction being the low-resolution data that is generated. For a detailed explanation of the theory of SAXS and the methods used to generate low-resolution structures of proteins see Chapter 2.2.1.

Circular Dichroism Spectroscopy

In some cases where high-resolution methods such as X-ray diffraction are not feasible, other techniques that yield less detailed results but still hold valuable information are used.

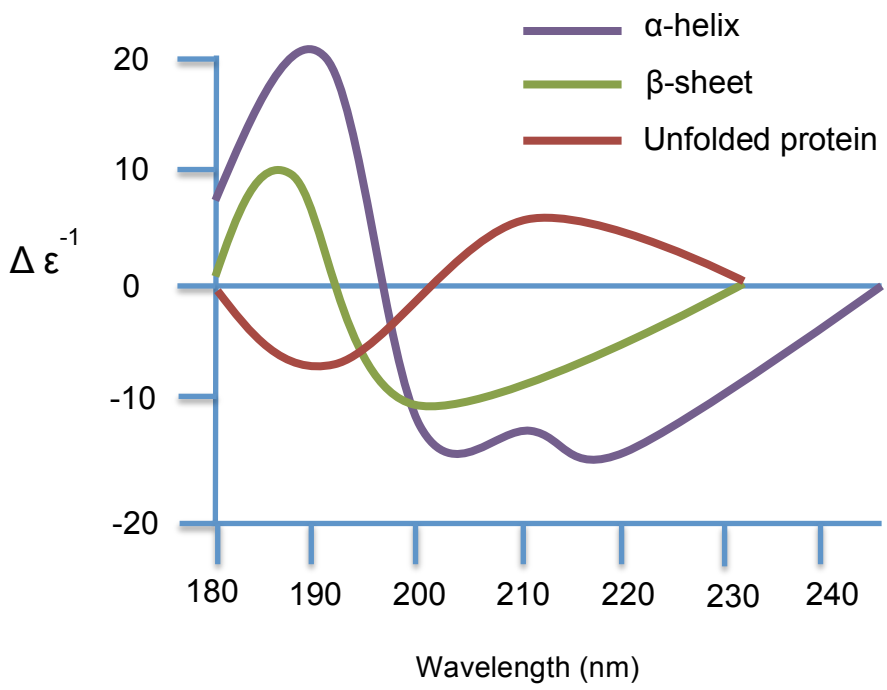


Figure 3.3.1a- CD spectra showing signatures for secondary structure. Adapted from Compton *et al*. (Compton and Johnson, 1986)

Unlike SAXS, which can detail overall shape, symmetry and domain arrangements in solution, circular dichroism (CD) spectra can provide useful information about the secondary structure of proteins. CD spectroscopy involves measuring the CD of a protein molecule over a range of wavelengths. CD occurs as a consequence of the interaction of polarised light with chiral molecules such as amino acids, and is defined as the difference between the absorption of left-handed circularly polarized light (L-CPL) and right-handed circularly polarised light (R-CPL)(Compton and Johnson, 1986; van Stokkum *et al.*, 1990).

CD spectroscopy is a useful technique that only requires small volumes of protein. The CD signatures can provide information about the α-helical and β-sheet secondary structural elements, see Figure 3.3.1a. There are various algorithms used to decipher the CD spectra signature, and in some such cases the percentage of each structural element in the protein can be predicted. CD spectroscopy will be used in this chapter to identify secondary structural features of the CDTb variants.

Size-Exclusion Chromatography Multi-Angle Light Scattering

As there is some ambiguity as to what the oligomeric state is of the CDTb variants as noted during size-exclusion chromatography, DLS can be used to determine whether or not the protein is monodisperse and can provide an approximate particle size. However, for more specific molar mass determination, size-exclusion chromatography multi-angle light scattering (SEC-MALS) can be employed.

SEC-MALS is a technique that combines a number of protein analysis techniques in one system. A size-exclusion column is linked to an HPLC machine with two additional detectors: a light scattering (LS) diode array and a refractive index (dRI) detector, see Figure 3.3.1b. This online measurement allows for calculation of the absolute molecular mass of each molecule eluted from the size-exclusion column. The light scattering module is similar to DLS as discussed in detail in Chapter 2.2.1, however unlike the DLS data collected in

Chapter 2.2.2 and Chapter 3.3.2, this module uses a multi-angle detector rather than a single angle detector as in DLS. When the electric field of a monochromatic coherent laser impacts a molecule in solution, a dipole is generated causing light scattering (Wang *et al.*, 1979). The intensity of scattered light is determined with a detector array at multiple angles, and this is directly proportional to the molar mass and concentration of the molecule.

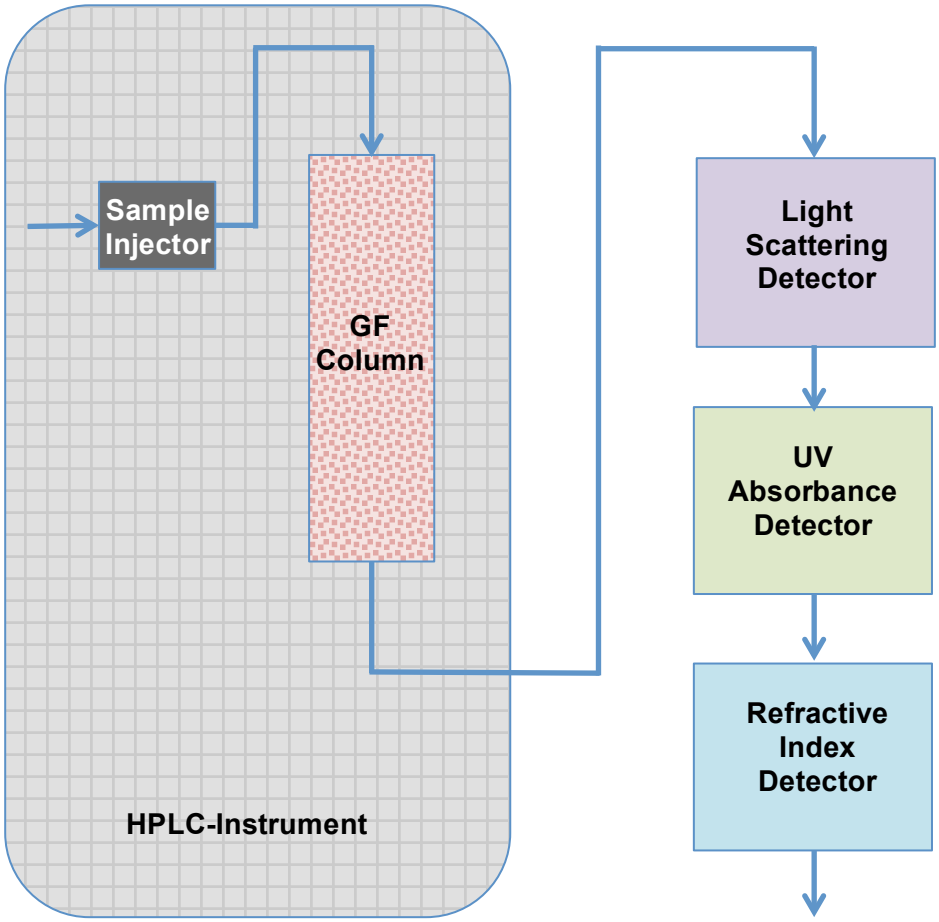


Figure 3.3.1b- A schematic diagram showing the SEC-MALS experimental set up.

The concentration of the molecules in solution is analysed by measuring the refractive index of the solution (dRI). The ratio of the LS to the dRI provides the absolute molar mass (Wang *et al.*, 1979; Wyatt, 1998). SEC-MALS will be used in this chapter to determine the absolute molar mass of CDTb' and CDTb'' which will resolve the current ambiguity of oligomeric state of both proteins.

Electron Microscopy

Negative stain electron microscopy is slowly becoming a popular method for low-resolution structure determination for proteins that are recalcitrant to crystallisation. This technique is most commonly used for membrane proteins, macromolecules and macromolecular complexes.

In this subchapter, the aim was to visualise CDTb'' using electron microscopy and SAXS to determine whether or not CDTb'' is forming the heptameric pore as predicted.

3.3.2 Materials and Methods

Dynamic Light Scattering

The various forms of CDTb were purified as discussed previously in Chapter 3.1.2. Due to time limitations, only the CDTb' and CDTb'' forms were prepared for DLS and SAXS. The protein samples were prepared by centrifugation at 13,000 rpm to remove any large particles or precipitate and were filtered through a 0.45 µm millex-HA filter (Millipore). Nitrile gloves were worn to avoid fingerprints on the sample cuvette and to prevent contamination of the samples during handling. A low volume 12 µl glass cuvette (QS 3.00 mm) was equilibrated by rinsing in the CDTb buffer: 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 mM DTT.

DLS was performed using a Zetasizer Nano S (Malvern Instruments) which uses a 4 mW He-Ne laser with wavelength λ_o = 633 nm, and a scattering angle of θ = 173°. A 50 µl sample of each of the CDTb forms at concentrations of 0.5 mg/ml were loaded in succession into the cuvette and the external surfaces of the cuvette were gently wiped with soft lens cleaning tissue to ensure no particles were present that could lead to false readings. The cuvette was placed in the sample holder and was allowed to equilibrate to 19 °C for 2 min. A series of three scans were performed for each CDTb form and subsequent intensity fluctuations were recorded which contained information about the time scale movement of the scattering particles. DLS results will confirm whether or not the CDTb samples are suitable for use in SAXS data collection.

Multiple crystallisation plates were set up in Chapter 3.2.3 after buffer exchanging CDTb variants into alternative buffer systems to provide a wider variety of crystallisation conditions. DLS data were also collected for the buffer exchanged CDTb' and CDTb'' variants to confirm that there was no effect on the protein such as degradation or aggregation that might interfere with crystal growth.

SAXS Data Collection and *Ab initio* Model Building

CDTb' and CDTb'' samples prepared as above were taken to Diamond Light Source for SAXS data collection. Experiments were performed as with Toxin A in Chapter 2.2.2, on the I22 beam line, at 20 °C using the Pilatus 2M detector. Concentrations of 0.5 mg/ml, 1 mg/ml, 2.5 mg/ml and 5 mg/ml were used in the scattering experiments, along with four buffer samples. Calibration was carried out using BSA at 5 mg/ml in 50 mM Tris-HCl pH 7.5 as a molecular weight standard. Data was integrated using Dream (in-house software, Diamond Light Source) and the subsequent data collected for the buffer samples was averaged and subtracted from the CDTb data using PRIMUS (Konarev *et al.*, 2003). An indirect Fourier Transformation automated by GNOM was used to generate a particle distribution function P(r) of the samples, which in turn was used to calculate the radius of gyration (R_g) and the maximum particle size (D_{max}) (Svergun, 1992). *Ab initio* modelling was performed using the on-line server for DAMMIF, whereby 20 models were produced (Franke and Svergun, 2009). The models were then averaged using DAMAVER (Volkov and Svergun, 2003). The final averaged *ab initio* model was then visualised using PyMOL (version 1.5.0.4, Schrödinger, LLC).

CD Spectroscopy

As discussed in Chapter 3.2.1, X-ray crystallography techniques have not been successful for the CDTb forms up until this point and so other techniques that yield less detailed results but still hold valuable information have been used. One such technique is CD Spectroscopy that provides secondary structural information of the protein.

CD spectroscopy was performed using a Jasco J-600 spectropolarimeter with a 2 mm rectangular cell. CDTb samples were diluted to 0.5 mg/ml and dispensed into the cell that had been pre-rinsed with CDTb buffer. The wavelength scans were performed at 20 ˚C over a range of 195-250 nm at a 50 nm/min rate.

Size-Exclusion Chromatography Multi-Angle Light Scattering

The SEC-MALS setup was as demonstrated in Figure 3.3.1b. A silica based size-exclusion column (TSK-G3000SW, Sigma Aldrich) was connected to an HPLC (Agilent Technologies, UK) instrument. This was then connected to a light scattering diode array (Dawn Heleos II, Wyatt, USA) and a differential refractive index detector (Optilab rEX, Wyatt, USA) in series. The column was pre-equilibrated overnight in CDTb'' buffer: 50 mM Tris-HCl pH 7.5 and 150 mM NaCl (2 mM DTT was removed to prevent any low-level interference with the UV absorbance detector) at 1 ml/min to achieve stable base lines as well as low background scatter.

Initially the final purification step of CDTb'' was performed during the SEC-MALS so that any monomer present would be detected and also GST as a control to ensure the size-exclusion column yielded similar results to those performed on the regular superdex-200 10/300 GL. 90 µl of the CDTb'' sample was injected onto the system. The respective molar masses of subsequent peaks were calculated from resulting LS and dRI data using the software ASTRA (Wyatt).

Concurrently a standard superdex-200 10/300 size-exclusion run was performed to separate the CDTb'' oligomer, monomer and GST using the standard purification technique as described in Chapter 3.1.2. 90 µl of the high molecular weight oligomeric peak was then injected into the SEC-MALS system. The predicted monomeric peak was also injected into the SEC-MALS system and also a concentrated sample of the monomeric peak to determine whether or not oligomerisation was affected by concentration.

All of the SEC-MALS experiments were performed at 1 ml/min in buffer 50 mM Tris-HCl pH 7.5 and 150 mM NaCl. The experiments were carried out with the help of Dr Christian Pernstich, in Professor Steve Halford's research laboratory, at the University of Bristol.

Electron Microscopy

CDTb'' at 4 mg/ml was centrifuged at 13,000 rpm to remove any large particles or precipitate and were filtered through a 0.45 µm millex-HA filter (Millipore). The prepared protein was then adsorbed onto glow discharged carbon-coated copper grids. The grids were then negatively stained using 0.75 % (w/v) uranyl formate. Micrographs were recorded at a magnification of 500,000x using a JEOL JEM1200EXII transmission electron microscope, operating at 120 kV. Electron micrographs were recorded with the help of Iain Harris and Ursula Potter at the University of Bath.

3.3.3 Results

Dynamic Light Scattering

Pure CDTb samples were diluted to 0.5 mg/ml for use in DLS experiments. Due to time restraints, DLS was performed only for CDTb' and CDTb'', at the range of pHs discussed in Chapter 3.2.2. Three DLS scans were performed consecutively at 19 °C for each sample and the Zetasizer software auto processed the scattering data to extract decay rates to produce a size distribution.

Figure 3.3.3a displays the DLS results for both CDTb' and CDTb'' in the standard purification buffer at pH 7.5. The correlation function for the collected data, is displayed in Figure 3.3.3a A2 and B2 for CDTb' and CDTb'' respectively. The red line represents the decay over time, which is extended and is slow suggesting a large particle size. In addition the baseline is flat indicating a lack of larger particles or aggregates. Figure 3.3.3a A3 and B3 displays the count rate for CDTb' and CDTb'' respectively, which represents the intensity fluctuations. Typically the smaller the particle the larger the fluctuations as the particles are moving faster in solution, whereas larger particles have lower fluctuations, such as those displayed in this figure. Finally, the intensity PSD graph displayed in Figure 3.3.3a A4 and B4 provides details on the intensity of scattered light from each particle and the size of the corresponding particle, for CDTb' and CDTb'' respectively. The X-axis shows the distribution of size classes whilst the Y-axis shows the relative intensity of scattered light. For both CDTb' and CDTb'' there is a single large intensity peak of about 16 % between 20-200 nm in diameter. This is to be expected as both CDTb' and CDTb'' behave in a similar manner during size exclusion and elute as if similar size proteins. However, if the theory of oligomerisation is correct then CDTb'' should appear approximately seven times bigger than CDTb', however that is not the case.

The Mie theory can be used to convert the fundamental size distribution generated from an intensity distribution to a volume distribution as shown in Figure 3.3.3a A1 and B1. Although the Mie theory assumes that the particles are

all spherical, homogenous and the optical properties of the particles are known, it can be useful to examine a volume PSD graph in comparison to an intensity PSD graph. The volume PSD describes the relative proportion of multiple components in the sample based on their mass or volume (rather than their scattering intensity). The single peaks on these graphs for CDTb' and CDTb'' suggest monodispersity for both proteins rendering them suitable for SAXS data collection (Wiscombe, 1980). The correlation function, count rate, intensity PSD and volume PSD graphs for CDTb' at pH 6.5 and 7.0 are displayed in Figure 3.3.3b and for pH 8.0 and 8.5 in Figure 3.3.3c. It is immediately apparent that there are multiple peaks on all of the intensity PSD graphs for CDTb' at all pH variations except for the original pH 7.5. However when looking at the relative proportion of these components based on the volume PSD graphs the sample appears to be predominantly monodisperse. This is reflected in the distribution results displayed in Table 3.3.3a, which will be discussed in detail later in this section.

The correlation function, count rate, intensity PSD and volume PSD graphs for CDTb'' at pH 6.5 and 7.0 are displayed in Figure 3.3.3d and for pH 8.0 and 8.5 in Figure 3.3.3e. At pH 6.5 and 7.0, the CDTb'' data does not look monodisperse, there are multiple peaks on the intensity PSD graphs and the same applies for the volume PSD graphs. The higher pH buffer; pH 8.5 doesn't appear to have affected CDTb'' dispersity as the results are similar to that of the standard pH 7.5 results. The pH 8.0 buffer however has affected the dispersity as reflected in the volume PSD graph which shows a slightly warped peak and a much larger peak which again could suggest the presence of aggregated protein or dust particles. In summary the DLS results indicate that pH 7.5 buffer is most suitable for CDTb' and CDTb'' to maintain the protein in a monodisperse state. All changes in pH affect the dispersity of CDTb' where as only the lower pH buffers have a significant affect on CDTb'' dispersity.

Table 3.3.3.a displays both the cumulants and distribution results obtained for CDTb' and CDTb'' for the different pH buffers. The Z-Average size is also known as the cumulants mean and is defined as the hydrodynamic particle size. This measurement represents the diameter of a sphere that has the same drag coefficient as a given particle, so the Z-Average size is a predicted particle size based on the movement of the standard particles in solution. This

estimated value is based on the theory that all of the particles are spherical in shape. The Z-Average particle size for CDTb' varies between 40.1 nm and 55.3 nm which is considered fairly consistent. This value indicates that CDTb' is approximately 400-550 Å, which is much larger than anticipated, especially considering this is the monomeric form of CDTb. However, this does correspond with the size-exclusion chromatogram whereby CDTb' elutes at a point expected for high molecular weight proteins > 500 kDa. The range of particle sizes for the differing pH buffers is broad enough to suggest that the pH of the buffer is altering the oligomeric state of the protein. The Z-average size particle diameter for CDTb'' ranges between 40.5 nm and 100.9 nm, which is a large range and indicates the protein could have an approximate diameter between 400-1000 Å. This range is far too broad which may indicate that the pH variation does affect the oligomeric arrangement of CDTb'' too.

The Polydispersity Index (PDI) is a measure of the width of molecular weight distributions, and is an indication of variance in the sample. A PDI value of 0.2 or less is expected for monodisperse samples, therefore from Table 3.3.3a it is clear that only two samples on the table are likely to be monodisperse, which are CDTb' and CDTb'' at pH 7.5. The remaining data all have PDI values of greater than 0.2, which indicates that these samples are not monodisperse.

The distribution result describes the percentage of each identified particles of the same size in the solution along with the estimated diameter size of each of these particles, corresponding to peaks on the intensity PSD graph. Both CDTb' and CDTb'' in the different pH buffers have multiple peaks with a variety of distribution results, but for the two samples mentioned above with low PDI values (CDTb' and CDTb'' at pH 7.5) the distribution result is 100 % and 99.7 % for CDTb' and CDTb'' respectively. The data collected at pH 7.5 indicates that both CDTb' and CDTb'' are approximately 530-560 Å and are monodisperse.

To conclude, all of the DLS results reveal that oligomeric states of CDTb' and CDTb'' vary with pH. The key finding is that CDTb' and CDTb'' share a similar overall particle size diameter of 530-560 Å at pH 7.5, which correlates with the size-exclusion chromatogram observed for both proteins.

Light Scattering Data Statistics		CDTb'					CDTb''				
		pH 6.5	**pH 7.0**	**pH 7.5**	**pH 8.0**	**pH 8.5**	**pH 6.5**	**pH 7.0**	**pH 7.5**	**pH 8.0**	**pH 8.5**
Cumulants result	Z-Average size	40.10 d.nm	55.29 d.nm	53.10 d.nm	54.15 d.nm	42.48 d.nm	78.44 d.nm	100.9 d.nm	56.57 d.nm	56.57 d.nm	40.52 d.nm
	PDI	0.432	0.453	0.182	0.725	0.444	0.290	0.529	0.187	0.372	0.229
Distribution result	Peak 1 (Mean/ Area)	13.70 d.nm/ 99.0%	10.94 d.nm/ 99.8 %	37.97 d.nm/ 100 %	12.02 d.nm/ 99.3 %	13.10 d.nm /99.5 %	29.63 d.nm/ 99.7%	79.17 d.nm/ 61.8%	42.51 d.nm/ 99.7 %	22.51 d.nm/ 97.5 %	28.15 d.nm/ 99.7 %
	Peak 2 (Mean/ Area)	3842 d.nm/ 0.1 %	3458 d.nm/ 0.2 %	-	443.3 d.nm/ 0.4 %	1934 d.nm/ 0.5 %	4456 d.nm/ 0.3 %	729.4 d.nm/ 36.9 %	4613 d.nm/ 0.3%	1831 d.nm/ 2.5 %	3932 d.nm/ 0.3 %
	Peak 3 (Mean/ Area)	-	-	-	3668 d.nm/ 0.3 %	-	-	5151 d.nm/ 1.3 %	-	-	-

Table 3.3.3a- Dynamic light scattering data statistics for both CDTb' and CDTb'' at a pH range of 6.5-8.5. The cumulants results include the Z-Average size, which is defined as the hydrodynamic particle size, and the Polydispersity Index (PDI) which is a measure of the molecular weight distributions. The distribution result describes the percentage of each identified particles of the same size in the solution and includes the estimated diameter size of these particles.

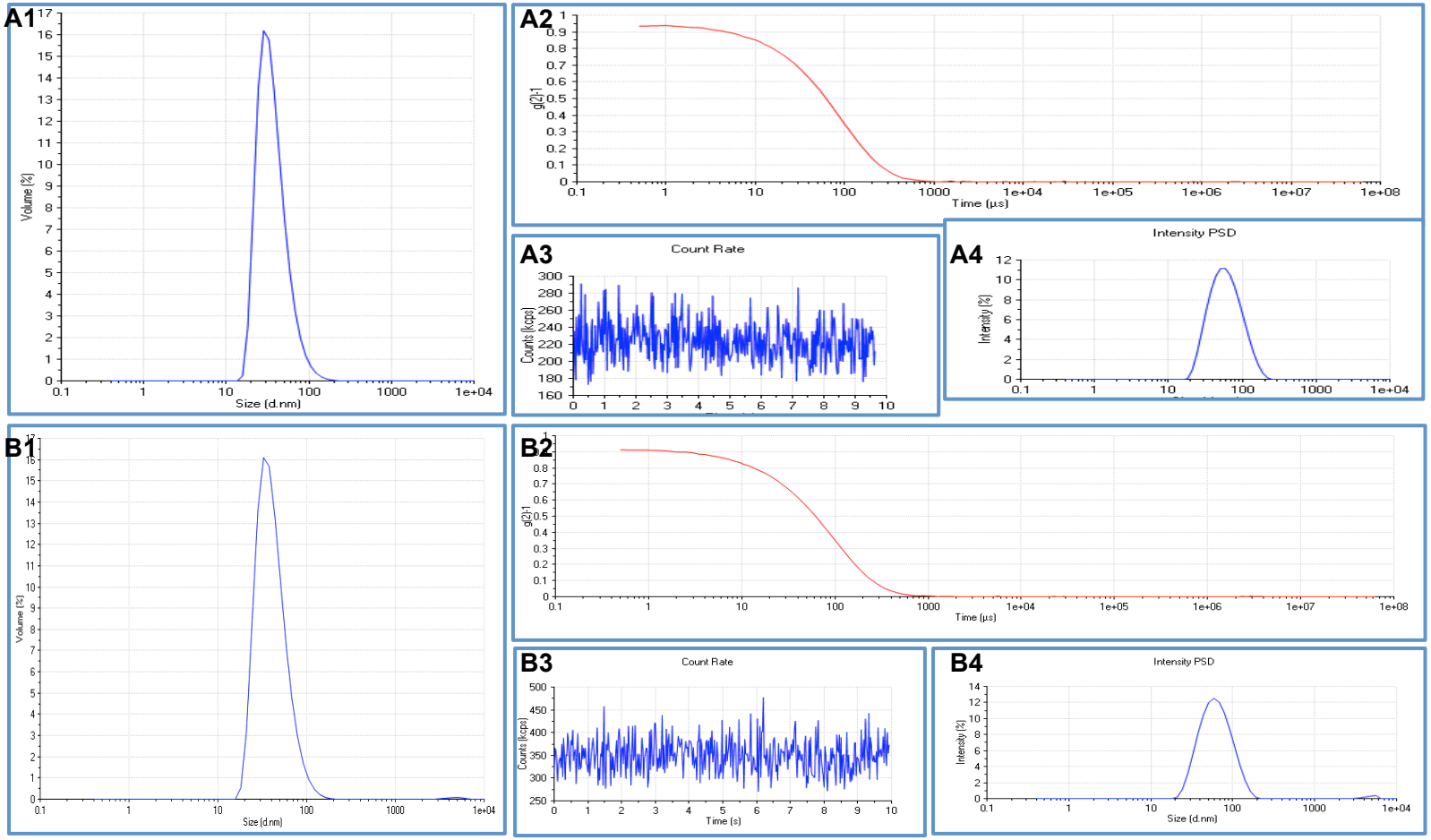


Figure 3.3.3a- Dynamic light scattering results for CDTb' at pH 7.5 (A) and CDTb'' at pH 7.5 (B). For both A and B: graph 1 displays the volume PSD, 2 displays the correlation function, 3 displays the count rate and 4 displays the intensity PSD.

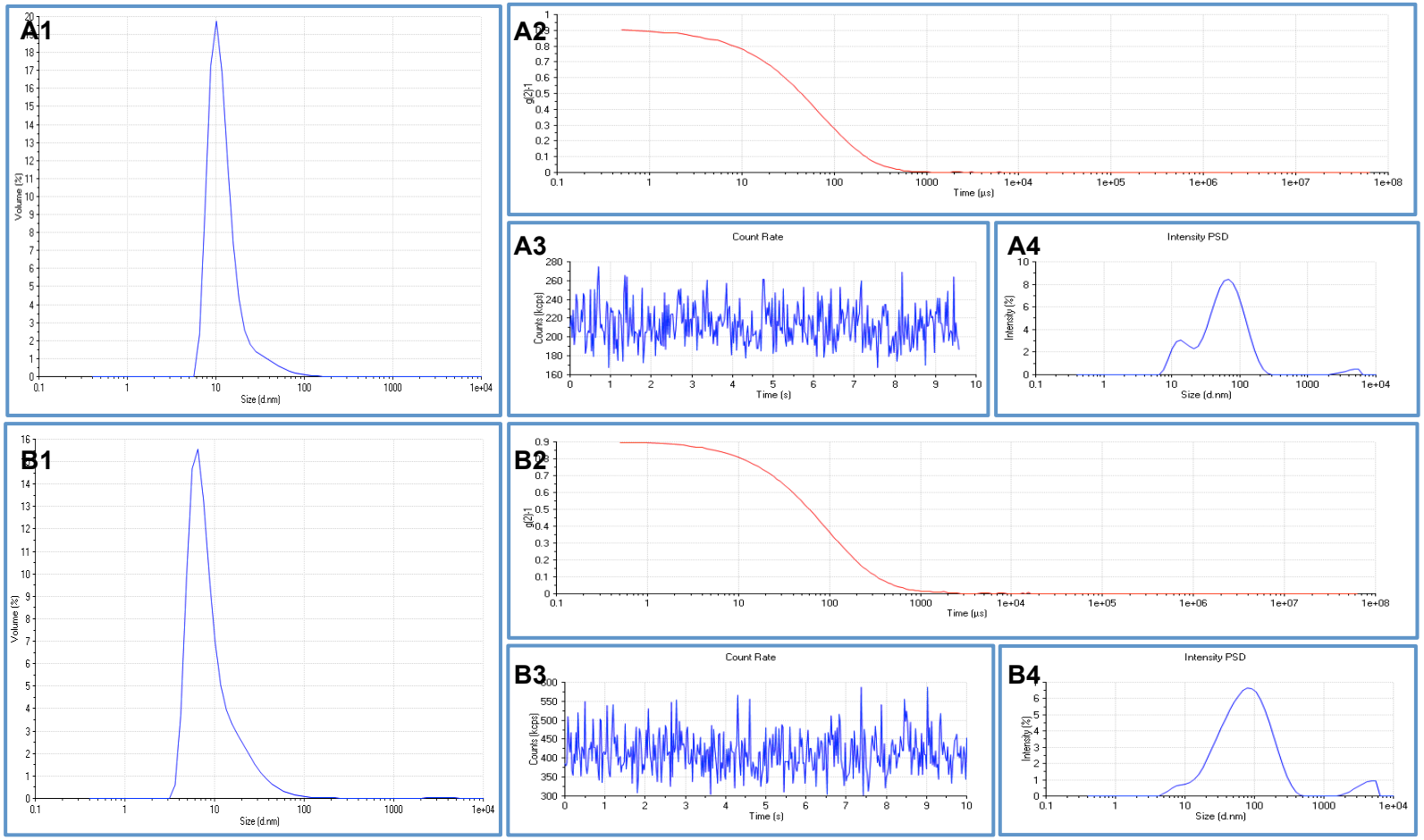


Figure 3.3.3b- Dynamic light scattering results for CDTb' at pH 6.5 (A) and pH 7.0 (B). **For both A and B: graph 1 displays the volume PSD, 2 displays the correlation function, 3 displays the count rate and 4 displays the intensity PSD.**

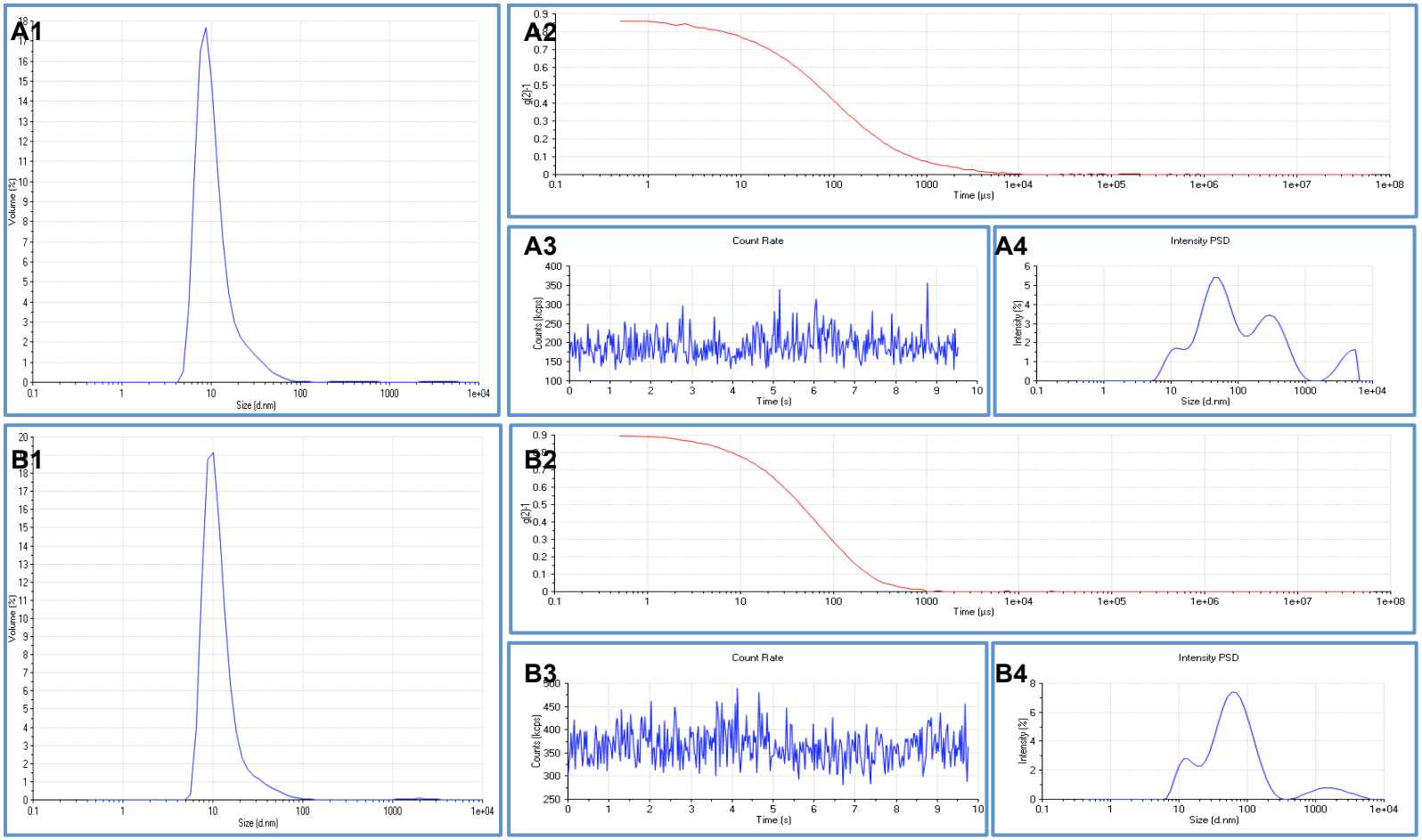


Figure 3.3.3c- Dynamic light scattering results for CDTb' at pH 8.0 (A) and pH 8.5 (B). **For both A and B: graph 1 displays the volume PSD, 2 displays the correlation function, 3 displays the count rate and 4 displays the intensity PSD.**

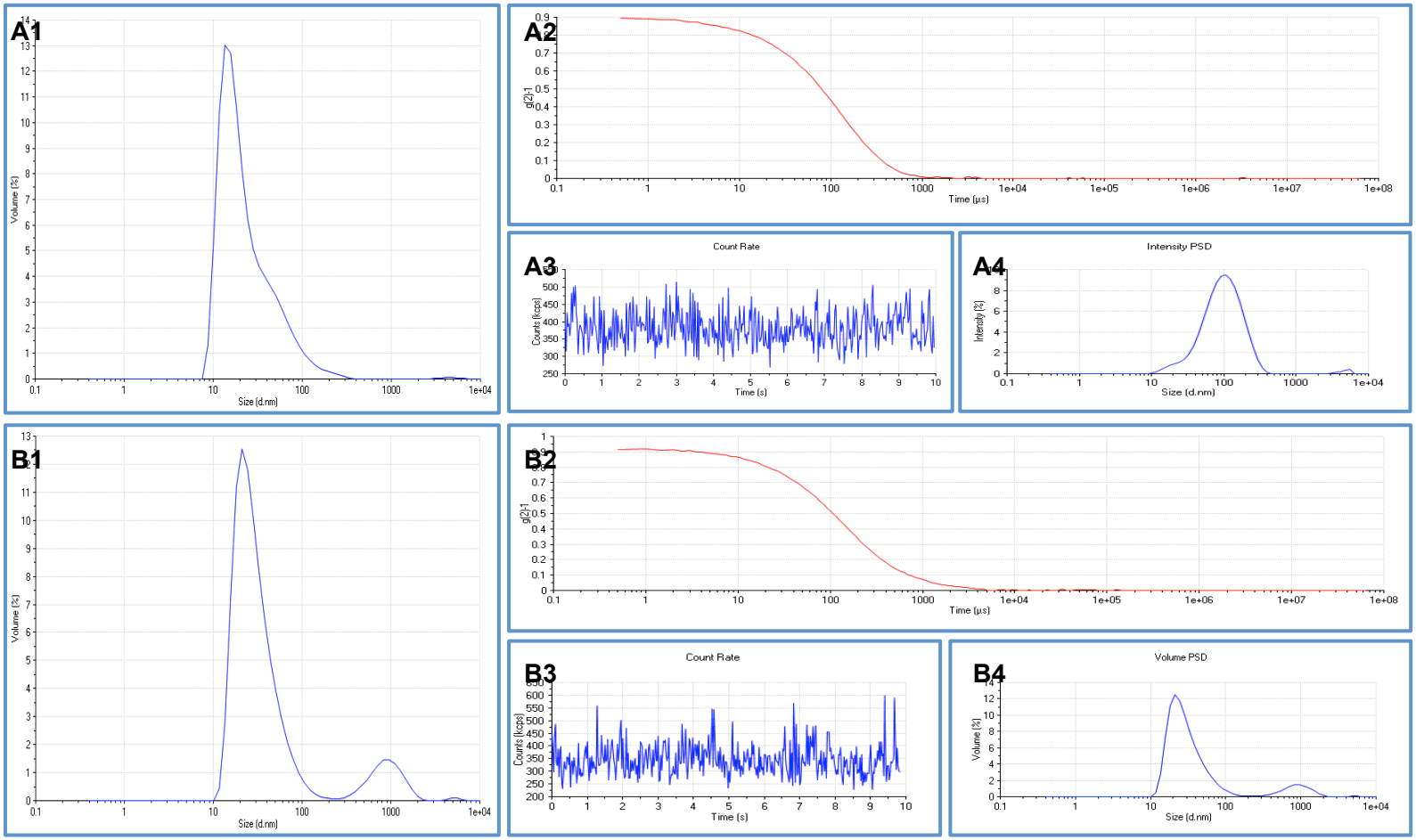


Figure 3.3.3d- Dynamic light scattering results for CDTb'' at pH 6.5 (A) and pH 7.0 (B). **For both A and B: graph 1 displays the volume PSD, 2 displays the correlation function, 3 displays the count rate and 4 displays the intensity PSD.**

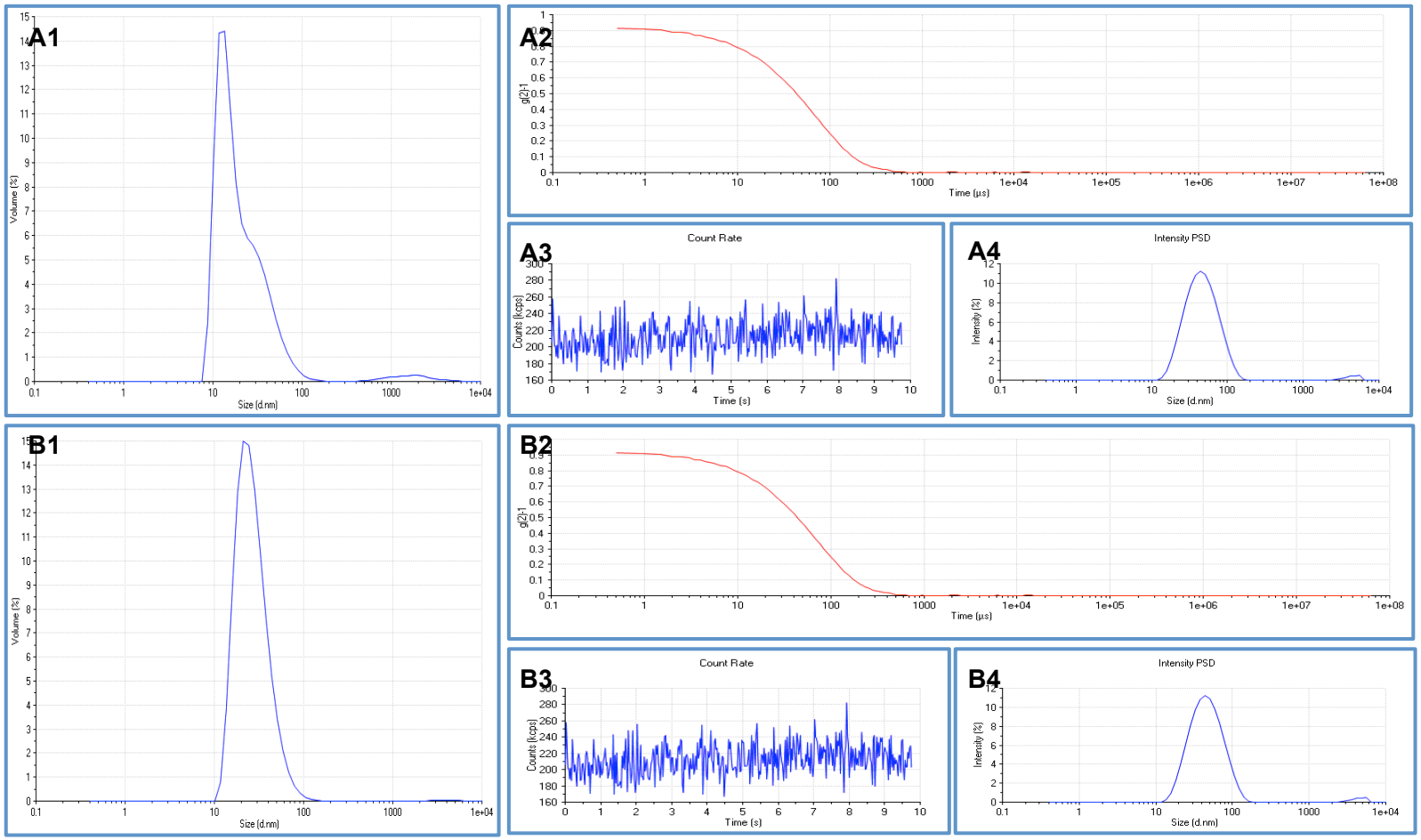


Figure 3.3.3e- Dynamic Light Scattering results for CDTb'' at pH 8.0 (A) and pH 8.5 (B). **For both A and B: graph 1 displays the volume PSD, 2 displays the correlation function, 3 displays the count rate and 4 displays the intensity PSD.**

Small-Angle X-Ray Scattering

CDTb' SAXS Structure

Given the absence of structural data for all CDTb variants, SAXS data were collected, processed and *ab initio* modelling was performed to generate a low resolution SAXS envelope. Initially the data was reduced and the solvent subtracted using PRIMUS (Konarev *et al.*, 2003). The scattering data displayed in Figure 3.3.3f-A was Fourier transformed using GNOM to obtain the P(r) function, also displayed in Figure 3.3.3f-B, from which the R_g was calculated.

Scattering Data Statistics		**CDTb'**
P(r) Function	R_g (Å)	139.6
	D_{max} (Å)	420
Structural Modeling	Discrepancy (χ^2)	0.038
	NSD	1.18 ± 0.02

Table 3.3.3b- Scattering data statistics obtained during SAXS data processing.

The R_g value was calculated using a Guinier plot, which is the analysis of the scattering curve at the smallest scattering angles. The R_g value describes the size of the particle in terms of the root mean square distance from the outer particle to its center. The R_g for CDTb' as stated in Table 3.3.3b was calculated to be 139.6 Å. The D_{max} value is the maximum particle dimension calculated from the P(r) curve in Figure 3.3.3f-B, with final result displayed in Table 3.3.3b. The D_{max} was calculated to be 420 Å, which correlates approximately to the Z-Average particle size obtained from DLS data. The profile of the P(r) curve can provide information as to the overall particle shape. In general, a bell-shaped curve indicates a perfect spherical shape. A shift in this bell-shaped curve to the left indicates a more elongated globular structure, which is indicated in the P(r) curve for CDTb' (Jacques and Trewhella, 2010).

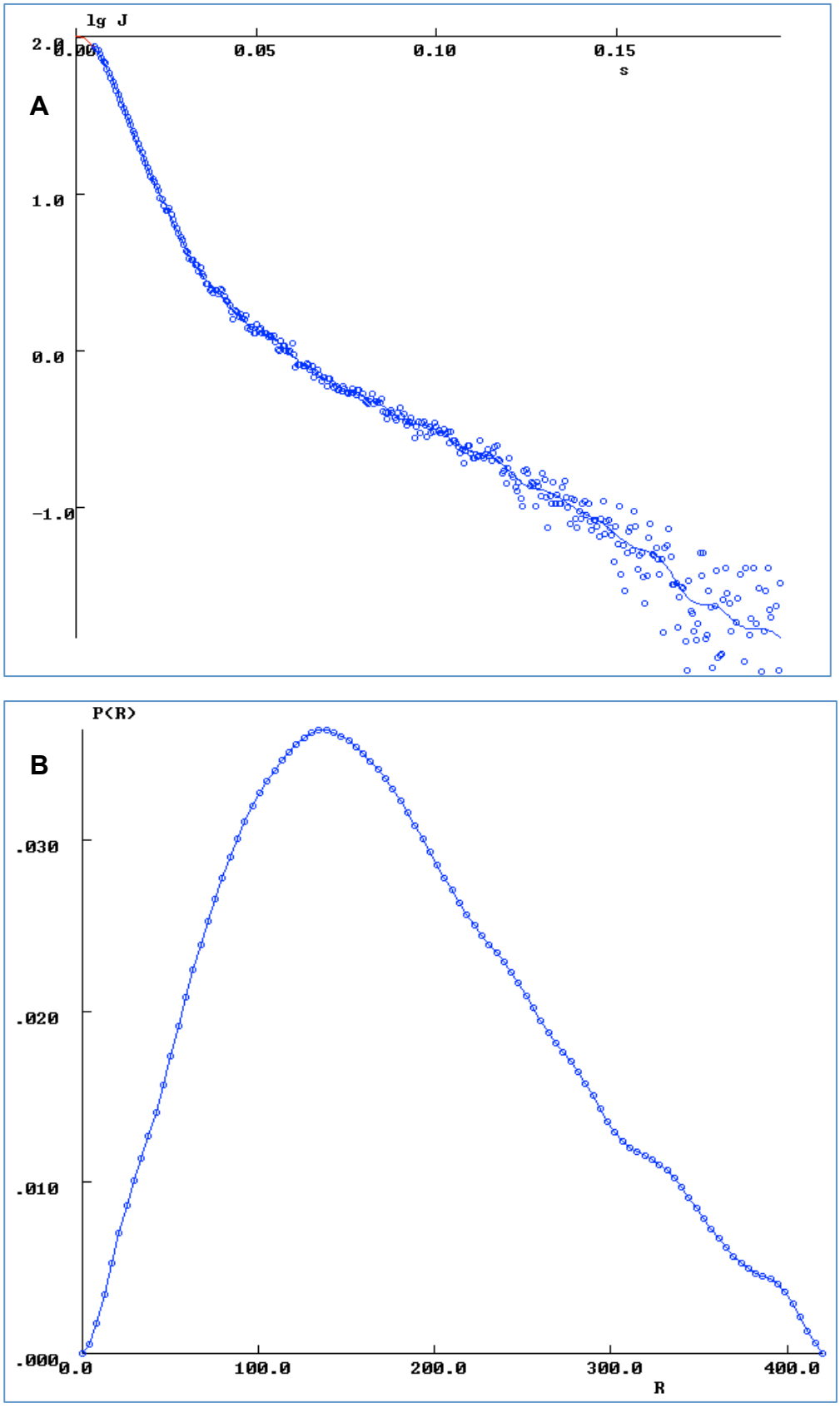


Figure 3.3.3f- Scattering curve and P(r) Curve for CDTb' (non-activated form of CDTb transport component). A: Scattering curve for CDTb' SAXS data. B: The P(r) curve indicates the SAXS envelope may have an elongated morphology.

DAMMIN was used to generate 20 *ab initio* models of CDTb' which were then averaged using DAMAVER to provide a final low-resolution SAXS envelope of CDTb' (Franke and Svergun, 2009; Volkov and Svergun, 2003). The structural modeling statistics are displayed in Table 3.3.3b. The discrepancy between experimental and *ab initio* data χ^2 was calculated using DAMMIF and the normalized special discrepancy (NSD) for the 20 models was calculated by DAMAVER (Franke and Svergun, 2009; Jacques and Trewhella, 2010; Volkov and Svergun, 2003). The χ^2 value is low suggesting low discrepancy between experimental and modeling data. In addition the low NSD value indicates there is little variation between the 20 models generated by DAMMIF.

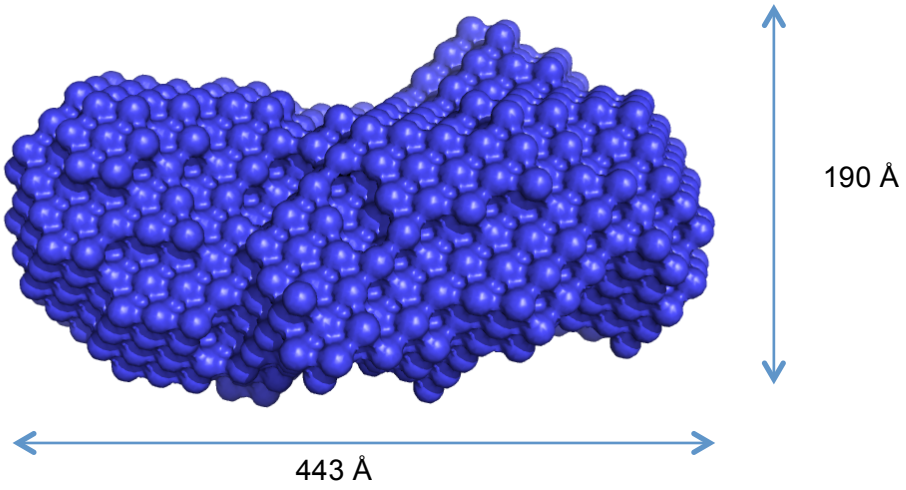


Figure 3.3.3g- *Ab initio* SAXS structure of CDTb'. The CDTb' SAXS envelope is approximately 443 Å in length and 190 Å wide, which is uncharacteristically large for a protein of 99 kDa. Images created using PyMOL (version 1.5.0.4, Schrödinger, LLC).

The final SAXS structure of CDTb' is displayed in Figure 3.3.3g, represented as a surface diagram in blue. The envelope is approximately 443 Å long and 190 Å wide. When you consider that CDTb' as a monomer is 99 kDa, this envelope is much larger than expected. A model of CDTb' was created using SWISS-MODEL with the C2 toxin from *Clostridium botulinum* as a model, see Figure 3.3.3h, which was then placed into the SAXS envelope of CDTb' for size comparison (Arnold *et al.*, 2006; Schleberger *et al.*, 2006). It is clear that the SAXS envelope is much larger than anticipated for a protein of this size.

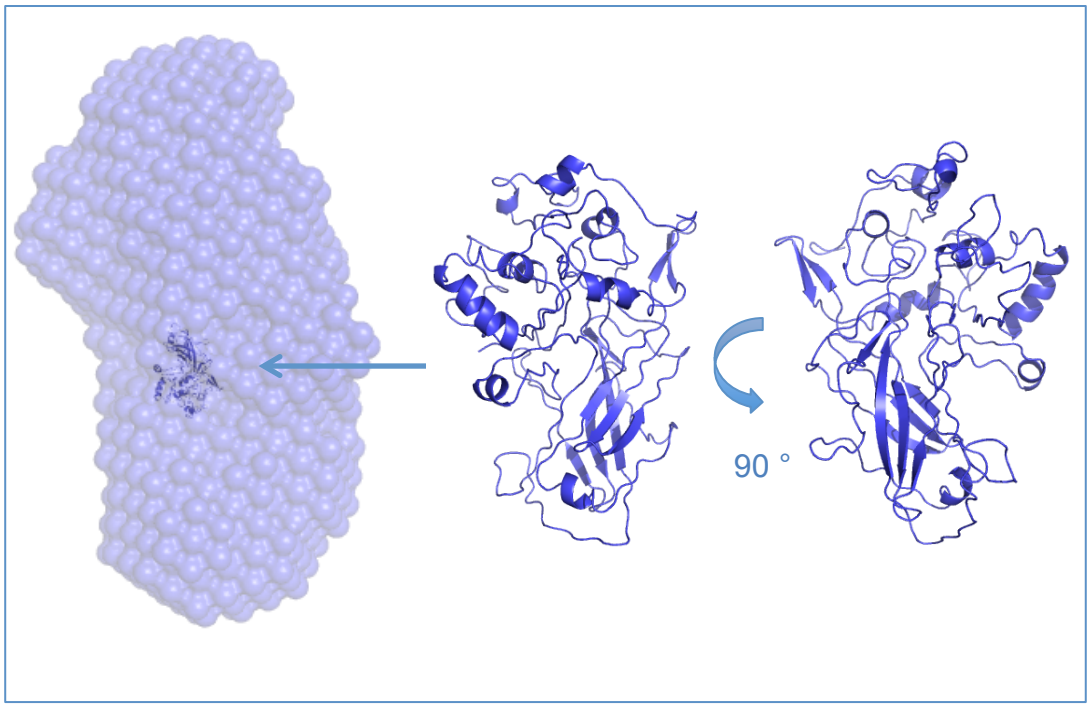


Figure 3.3.3h- *Ab initio* SAXS structure of CDTb' with CDTb' model placed in the SAXS envelope. The CDTb' SAXS envelope is approximately 443 Å in length and 190 Å wide, which is uncharacteristically large for a protein of 99 KDa. Displayed above is the CDTb' monomer model created using C2 toxin from *Clostridium botulinum*, superimposed into the SAXS envelope to illustrate the size differences (PDB: 2J42) (Schleberger *et al.*, 2006).

CDTb'' SAXS Structure

SAXS data for CDTb'' were collected in the same manner as CDTb', using the samples that had been analysed by DLS. On plotting of the scattering curve, it became apparent that the sample might have aggregated. A suitable Guinier plot could not be produced which caused inconsistent Rg calculations. The Guinier plot is taken from the scattering curve at the smallest scattering angle and inspection of the Guinier region can reveal the presence of aggregation in the sample which can result in non-reliable model generation (Putnam *et al.*, 2007). The data scattering at the smallest scattering angle was non-linear for CDTb'', hence the difficulties in calculation of an accurate R_g value. If the Guinier plot is non-linear, the protein sample may be aggregated and so data-processing was not taken further for CDTb''.

Circular Dichroism Spectroscopy

Circular Dichroism Spectroscopy was performed on CDTb' and CDTb'', primarily for secondary structure information, but also to cross check the results with the secondary structure features in the homology model created using C2 toxin from *C. botulinum*.

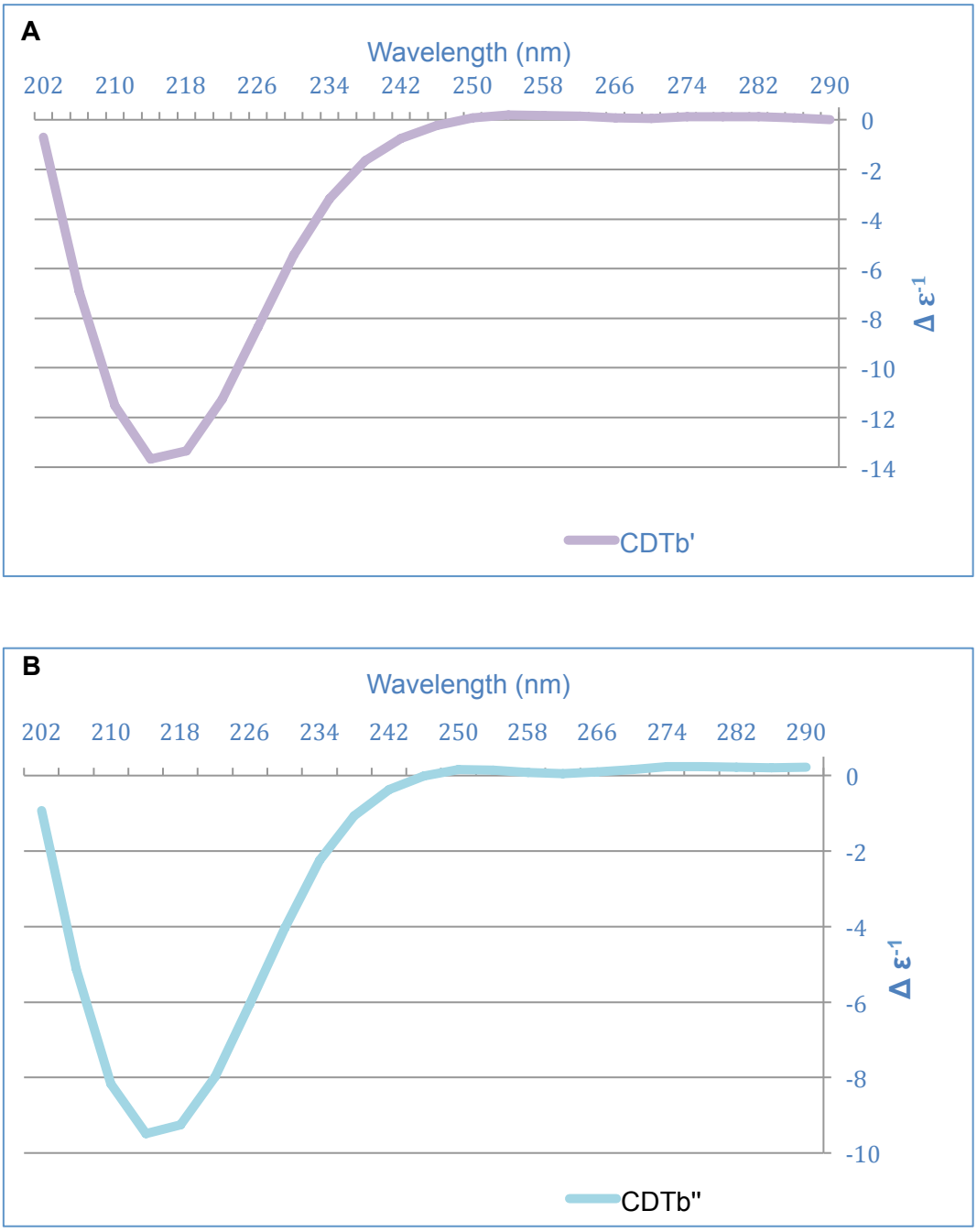


Figure 3.3.3i – Circular dichroism spectroscopy data for CDTb' (A) and CDTb'' (B).

Both the CDTb' and CDTb'' spectra share the same CD spectra profile. The $\Delta \varepsilon^{-1}$ starts to decrease at approximately 202 nm and reaches its minimum at approximately 217 nm whereupon the $\Delta \varepsilon^{-1}$ starts to increase until the base line is reached at 250 nm. Unlike the CD spectra from Figure 3.3.1a, the Jasco J-600 spectropolarimeter used for this experiments doesn't have the same wavelength range. The CDTb' and CDTb'' spectra profiles do not follow the signature profile usually observed for unfolded protein, which indicates both CDTb' and CDTb'' are folded. The CDTb' and CDTb'' spectra do share similarities to the signature profile for both α-helices and β-sheets, indicating that they have α-helical and β-sheet content. The percentage of α-helices and β-sheets in the samples cannot accurately be determined using data collected from this equipment due to the small wavelength range. The C2 binary toxin transport component contains both α-helices and β-sheets, which indicates the model may be a fairly accurate CDTb' representation.

Size-Exclusion Chromatography Multi-Angle Light Scattering

The SEC-MALS technique was used to determine the molar mass of CDTb''. The theory is that CDTb' is trypsin activated to form CDTb'' which then oligomerises, most likely to a heptamer form prior to CDTa binding. All of the CDTb variants elute from the gel-filtration column at the same point which is expected for proteins > 500 kDa in size. Initially it was thought that both CDTb and CDTb' could also form heptamers prior to activation. However it has slowly become apparent from both DLS data and SAXS data that the heptameric pore may not have formed, and it is more likely that the CDTb variants are in a different oligomeric state, potentially multiple states. SEC-MALS was performed using purified CDTb'' to determine the molar mass of the peak that elutes from the gel-filtration column at approximately 500 kDa. A crude sample of CDTb'' that had not been subjected to the final size-exclusion chromatography cleaning step was injected onto the SEC-MALS system, so that the CDTb' monomeric peak and GST peak would be present for control comparisons of molar mass calculations. In addition, some of the crude CDTb'' was subjected to the final cleaning step and the monomeric peak eluted at approximately 15 ml was injected onto the SEC-MALS system on a separate run so that the results could be directly compared, see Figure 3.3.3j.

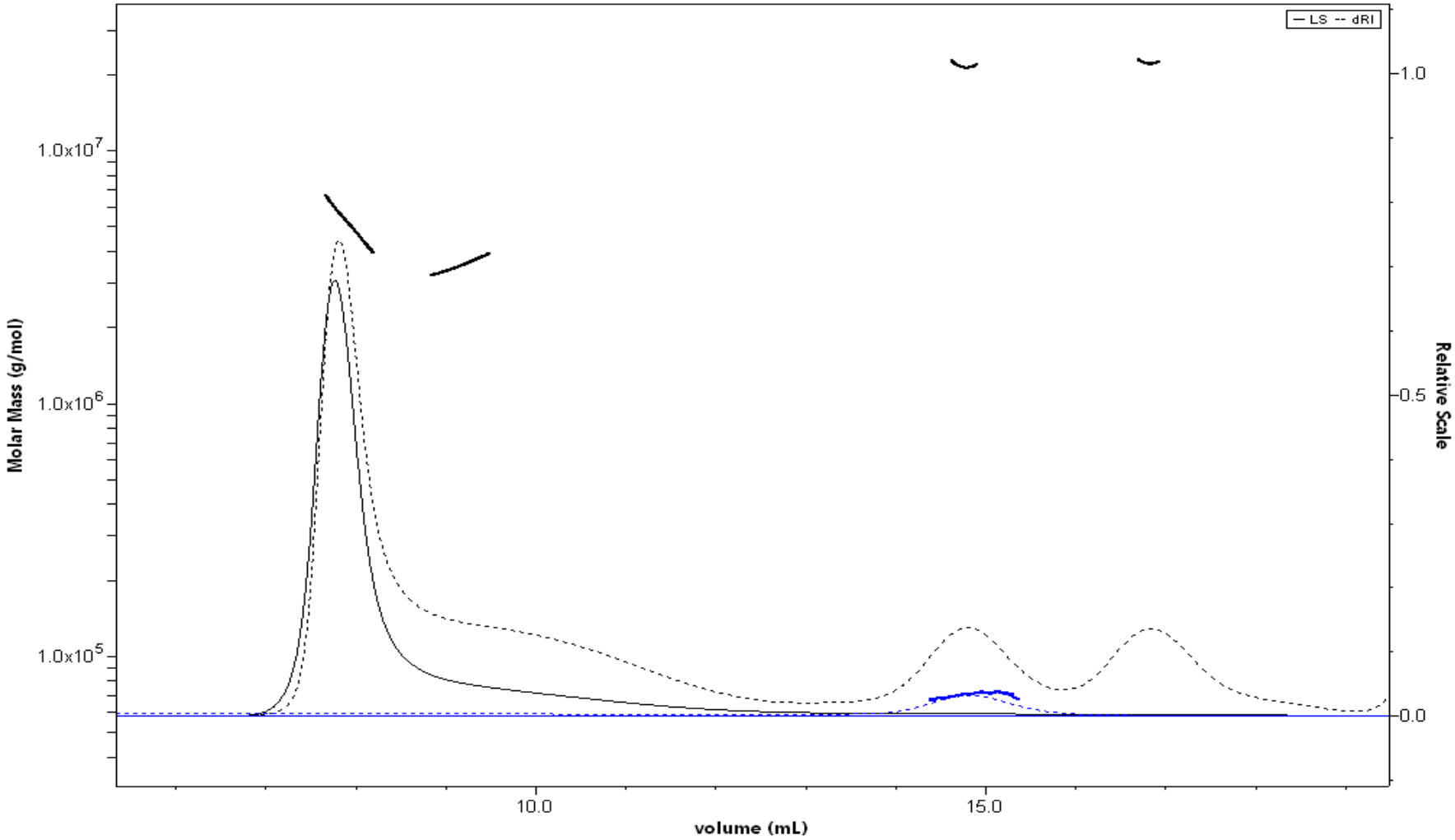


Figure 3.3.3j- SEC-MALS CDTb'' crude fraction results (black line) superimposed onto the SEC-MALS CDTb'' monomeric peak results (blue line). Note the monomeric peak usually elutes at approximately 15 ml and the oligomeric peak at approximately 7 ml. The lines above the peaks represent the molar mass calculations for the protein eluted in those fractions. Dotted line represents dRI and continuous line represents LS

Figure 3.3.3j displays the CDTb'' crude SEC-MALS results (black line) superimposed on the CDTb'' monomeric peak fraction results (blue line). The continuous line represents the Rayleigh Ratio (LS) and the dotted line represents the differential refractive index trace (dRI). The Rayleigh Ratio is an indication of the size and weight of the particle whereas the dRI is a measurement of concentration comparable to a UV trace.

For the crude CDTb'' sample, the graph displays the standard profile observed in the final cleaning step, with the oligomeric peak eluting at 7-8 ml, the monomeric peak eluting at 15 ml and the GST elution peak shortly after. The oligomeric peak dRI and LS are much higher than that for the monomeric peak which is consistent with the cleaning step chromatogram profile. The horizontal line across each peak indicates the selected data used for the molecular mass calculations performed using the software ASTRA (Wyatt). A reasonably flat line indicates a mono-dispersed species whereas a rapidly descending or ascending line is indicative of dynamically interchanging species. The oligomeric peak shows a rapidly descending line indicating this fraction of CDTb'' is actually multiple conformations of species rather than a specific oligomeric species as originally predicted. The CDTb'' monomeric peak at 15 ml shows a relatively flat line of approximately 0.7×10^5 g/mol which is 70 kDa. CDTb'' in monomeric form is 67 kDa so this SEC-MALS results confirms the peak eluting at 15 ml is CDTb'' in monomeric form. The CDTb'' oligomeric peak has a calculated molar mass range of 1.3×10^6 g/mol to 1.7×10^6 g/mol which is 1.3-1.7 MDa.

A sample of BSA was also injected onto the SEC-MALS system as a control. The calculated molar mass of BSA from the SEC-MALS data is displayed in Figure 3.3.3k in pink at approximately 60 kDa. The exactly molar mass of BSA is 66 kDa, so this would indicate the system is working properly. This result was superimposed onto the CDTb'' monomeric fraction result which has a calculated molar mass from the SEC-MALS of slightly under 70 kDa which is expected for CDTb'' given the specific molar mass is 67 kDa.

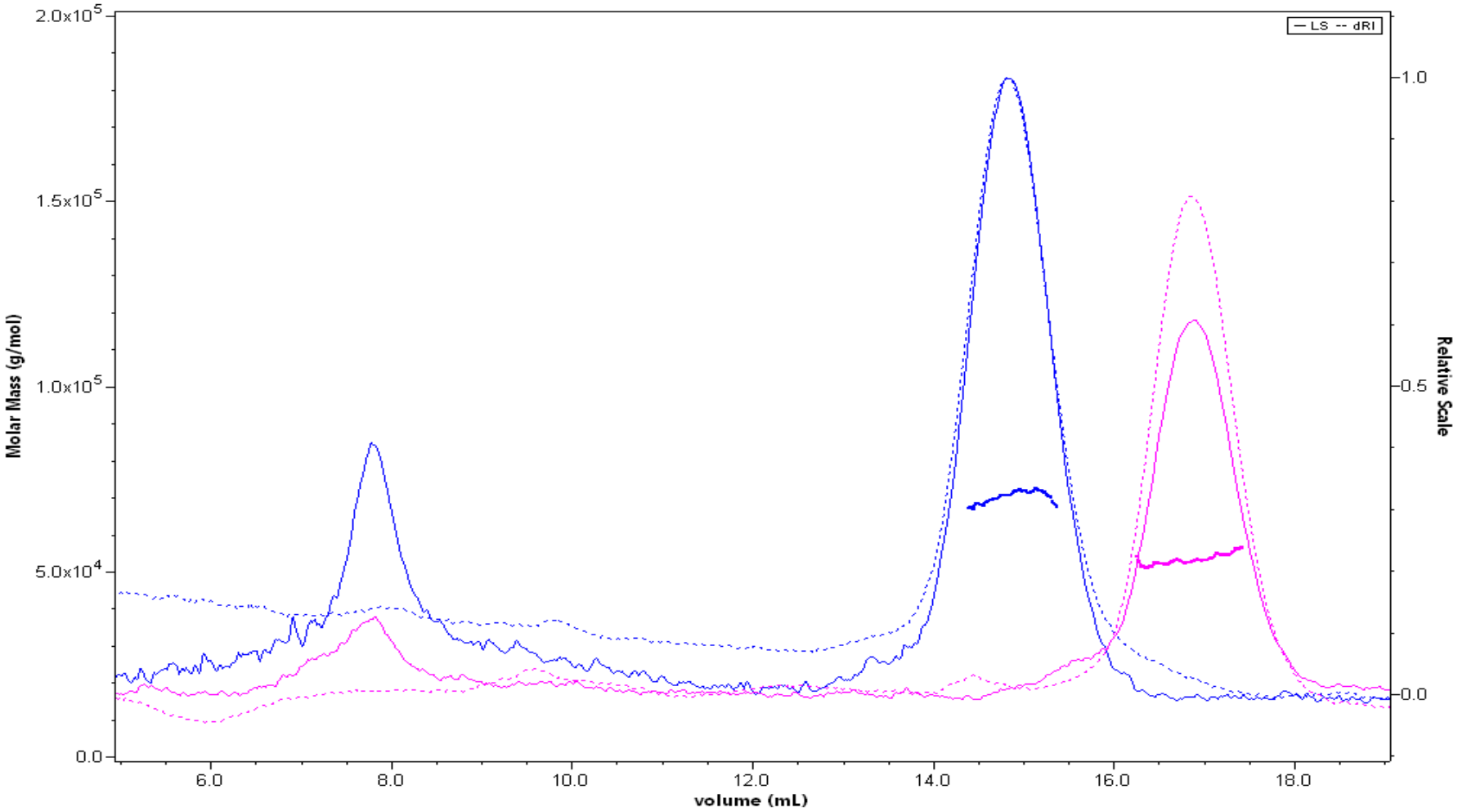


Figure 3.3.3k- SEC-MALS CDTb'' monomeric peak fraction results (blue line) superimposed onto the SEC-MALS BSA control results (pink). Dotted line represents dRI and continuous line represents LS. The lines within the peaks represent the molar mass calculations for the protein eluted in those fractions.

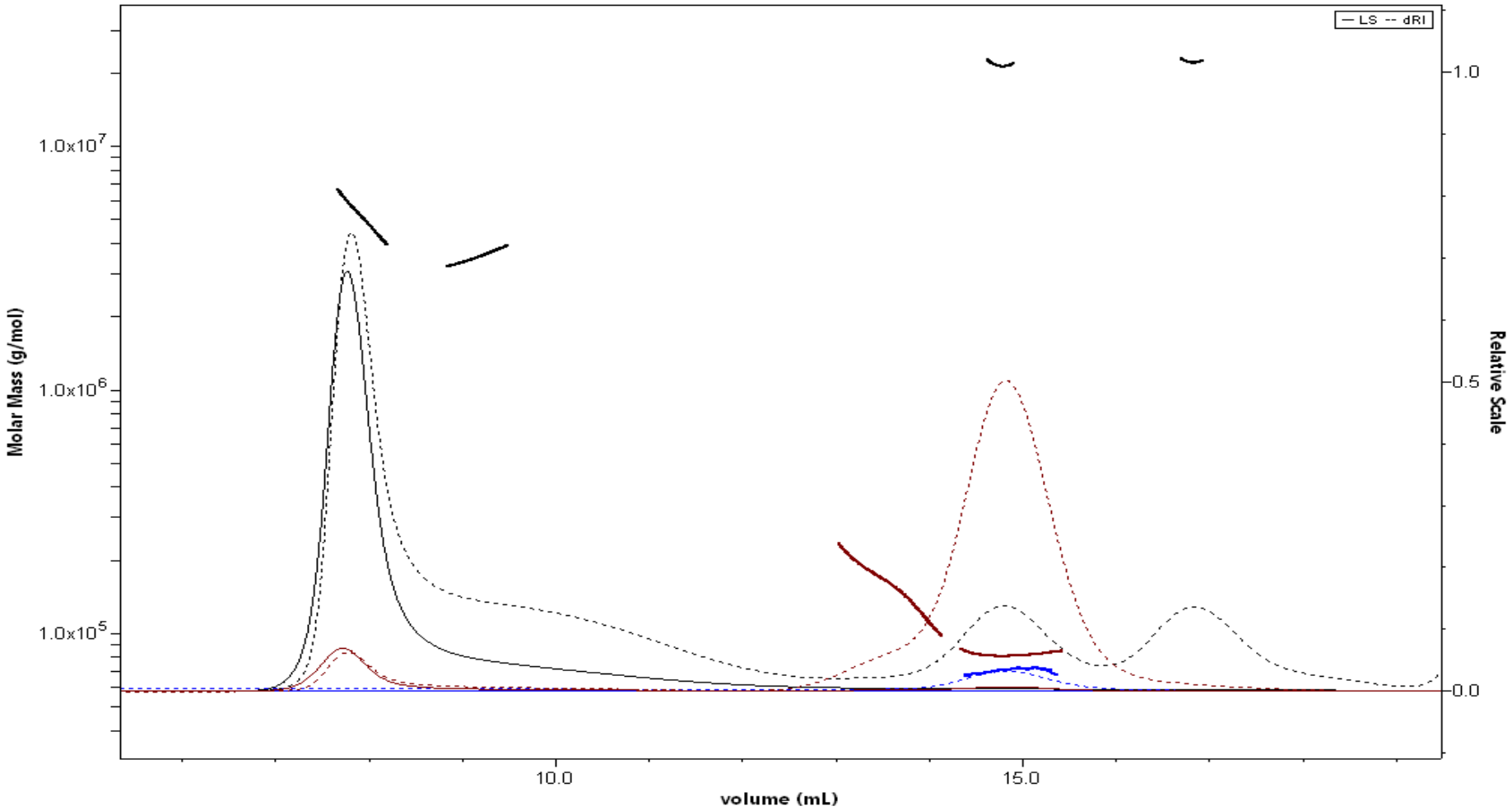


Figure 3.3.3l- SEC-MALS CDTb'' monomeric peak fraction results (blue line) superimposed onto the SEC-MALS CDTb'' crude results (black) and the concentrated CDTb'' monomeric peak in brown. **Dotted line represents dRI and continuous line represents LS. The lines above the peaks represent the molar mass calculations for the protein eluted in those fractions.**

Figure 3.3.3l displays the SEC-MALS result for the original crude sample (black), the CDTb'' monomeric fraction (blue) and the CDTb'' monomeric fraction concentrated up (red), to determine whether or not oligomerisation into dynamically interchanging species is a result of concentration. The key observation from this graph is that after concentration of CDTb'' monomer a peak has appeared at the same position observed for oligomeric CDTb'' that was not there previously at lower CDTb'' monomer concentrations indicating oligomerisation is observed at high concentrations but not low concentrations.

Electron Microscopy for CDTb''

CDTb'' was purified using the standard optimised protocol as discussed in Chapter 3.1.3 and was adsorbed to a glow-discharged carbon-coated copper grid. The grid was then stained with 0.75 % uranyl formate and was visualised using a JEOL JEM1200EXII transmission electron microscope. The results are displayed in Figure 3.3.3m.

Figure 3.3.3m displays the electron micrographs of CDTb'' at 4 mg/ml. Image A from Figure 3.3.3m displays the first electron micrograph of CDTb'' that was taken at the very edge of the copper grid. It is clear from this that the concentration used was too high and in future a concentration of 0.5 mg/ml may be more suitable.

The electron micrographs reveal that CDTb'' presents as large globular spheres. They are not clear heptameric pore-shaped as would be expected based on the theory of CDTb'' oligomerisation, but large inconsistently shaped spheres. A closer look as displayed in Figure 3.3.3m B shows that they do differ in size and are not perfectly spherical. They range from approximately 3 nm to 10 nm, which is consistent with results from DLS. Another distinctive feature of the electron micrographs is the presence of a variety of large well-structured species, Figure 3.3.3m-C and D. Some of these structures are elongated and with rigid edges whereas some are more irregular in shape but again with rigid edges. It is possible these are CDTb'' microcrystals.

.

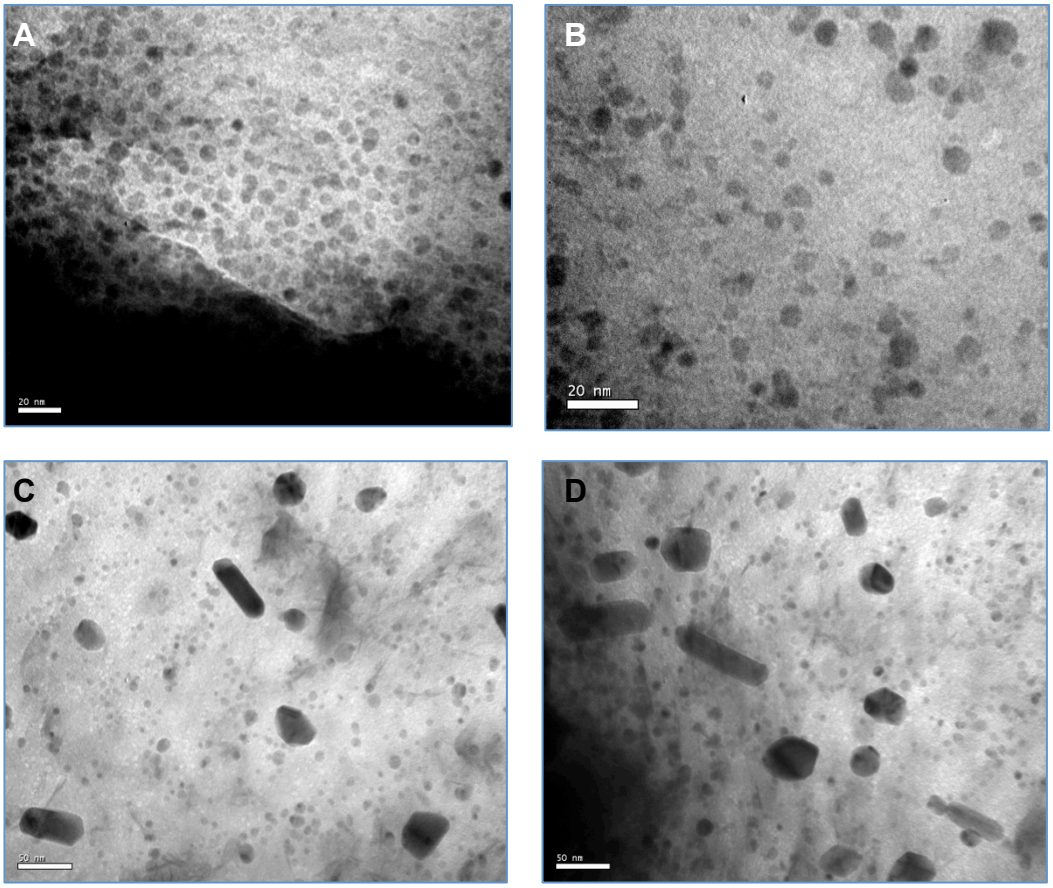


Figure 3.3.3m- Electron micrographs of CDTb''. All images are taken from the same plate. A: Scale bar represents 20 nm. B: A zoomed in version with scale bar representing 20 nm. C: Scale bar represents 50 nm. D: scale bar represents 50 nm.

3.4 Discussion

CDTb Variant Expression, Purification and Crystallisation

Some *Clostridium difficile* strains produce an ADP-ribosylating binary toxin; CDT, alongside the two potent exotoxins; Toxin A and B. The role of the binary toxin in pathogenicity is unclear, however some studies have shown that CDT causes the formation of microtubule-based protrusions on the cell surface of target cells (Schwan *et al.*, 2009). It is postulated that these protrusions promote adherence of *C.difficile* to the intestinal epithelial cells (Schwan *et al.*, 2009). The binary toxin is made up of two independently produced components; CDTa, the enzymatic component and CDTb, the transport component. The transport component is thought to interact with cell surface receptors on target cells. This component forms oligomers, interacts with the enzymatic component, CDTa, and both components are taken up into the cells by receptor mediated endocytosis. Studies have shown that individually CDTa and CDTb are non-toxic, however when the components are combined CDT is toxic to Vero cells (Sundriyal *et al.*, 2010).

CDTb is produced independently from CDTa and has an N-terminal signal peptide and an activation domain, see Figure 3.1.1a for the domain organisation. It is thought that CDTb must be activated by trypsin cleavage to form the full matured CDTb'' form. The aim of this chapter was to crystallise the individual CDTb variants: CDTb, CDTb' and CDTb'' and solve their structures using X-ray crystallography. Elucidating the high-resolution structures of the individual CDTb variants may contribute towards a greater understanding of the process by which CDTb oligomerises in preparation for translocation into the cytosol.

CDTb, CDTb' and CDTb'' were all expressed and purified successfully using the optimised technique as outlined in Chapter 3.1.2. One surprising observation made during the final cleaning step of size-exclusion was that all three variants eluted from the column at a point which is expected for proteins of > 500 kDa. This was unexpected as it is thought that only CDTb'' forms

oligomers, however, it was assumed that potentially all three variants form oligomers and so crystallisation trials were initiated.

Extensive attempts were made at crystallisation of CDTb variants, with some success. Diffraction quality crystals were obtained for CDTb' and CDTb'' but crystals stopped diffracting shortly after exposure to the beam. It is possible that this peak shift to high molecular weight observed at the size-exclusion chromatography step could be affecting crystallisation of CDTb variants, as obtaining crystals of CDTb, CDTb' and CDTb'' has proved extremely difficult. Therefore, efforts were made to elucidate the oligomeric state of the proteins.

Alternative Structural and Biochemical Techniques

Dynamic Light Scattering

Due to time restraints, CDTb' and CDTb'' were chosen for oligomeric state determination as these are functionally the most important CDTb variants. Initially DLS was used to determine whether or not the eluted protein was monodisperse because the superdex-200 10/300 column used in the final size-exclusion step cannot resolve proteins > 500 kDa, so it is possible that the peak observed contains multiple oligomeric arrangements. There are some homologous pore-forming toxins that are known to form more than one type of oligomer, for example the anthrax protective antigen (PA) transport component of the anthrax binary toxin (Atx) forms both heptamers and octamers (Kintzer *et al.*, 2009). The DLS data suggests that in the standard Tris-HCl pH 7.5 buffer used to purify CDTb' and CDTb'', the proteins are 100 % and 99.7 % monodisperse respectively. This is a strong indication that both CDTb' and CDTb'' are forming single state homogenous oligomers. The Z-average particle size prediction for CDTb' and CDTb'' from the DLS data processing are 53.1 nm and 56.6 nm respectively. As CDTb' is slightly larger at 94 kDa than CDTb'' at 67 kDa, but with a lower predicted Z-average size, it is possible that the variants are forming different sized oligomers. As discussed in Chapter 3.3.3, changing the pH of the buffer appears to affect the dispersity of the proteins, therefore pH 7.5 was kept consistent for the remaining alternative techniques used to characterise CDTb variants.

Small-Angle X-Ray Scattering

The SAXS data processing results are displayed in Chapter 3.3.3 along with the *ab initio* model generated for CDTb'. There were some difficulties encountered during processing of the data for CDTb', as there were some indications that the samples may be aggregated; however the DLS data suggests otherwise. There were problems generating a liner Guinier plot, however there were consistencies in calculation of R_g values and so the data were taken to the model building stage, with the final SAXS structure displayed in Figure 3.3.3g. The structure is approximately 443 Å long and 190 Å wide, which is slightly smaller than expected when compared to the predicted Z-average size from the DLS data of 530 Å. During processing of CDTb'' SAXS data it became apparent from the Guinier plot that the CDTb'' samples may be significantly aggregated, which again does not correspond to the DLS results, therefore data were not processed further so as to not cause biased model building.

Size-Exclusion Chromatography Multi-Angle Light Scattering

As the DLS data and SAXS data are creating some ambiguity as to the oligomeric state of CDTb' and CDTb'', SEC-MALS experiments were performed to calculate the exact molar mass of the high molecular weight peak eluted from the size-exclusion column in the final purification step of CDTb''. SEC-MALS combines a size-exclusion column with a UV detector, light scattering (LS) detector and a refractive index detector (dRI) (Wang *et al.*, 1979; Wyatt, 1998). The ratio of LS to dRI provides the absolute molar mass of a particular selected peak. The SEC-MALS results are displayed in Figures 3.3.3j, 3.3.3k and 3.3.3l. SEC-MALS confirmed that the high molecular weight elution peak from the size-exclusion column is in fact a mixture of dynamically interchanging oligomeric species in the mega Dalton range of sizes. The second peak observed in the final purification step was confirmed as monomeric CDTb'', which was noted to be of very low concentration. This fraction was then concentrated and again analysed by SEC-MALS. After concentration, this peak shifted to the left joining the high molecular weight fraction forming interchanging oligomers. Therefore, this confirms that oligomerisation occurs as a result of concentration. This may

explain why the DLS results suggest monodispersity as they are performed at 0.5 mg/ml whereas the SAXS data was collected at a range of 0.5-5.0 mg/ml.

Electron Microscopy

CDTb'' was concentrated to 4 mg/ml and adsorbed onto to a glow-discharged carbon-coated copper grid, stained with 0.75 % uranyl formate. A concentrated sample of CDTb'' was used so that these interchangeable oligomers could be visualised directly at low-resolution. The electron micrographs displayed an array of large inconsistently shaped species, which was to be expected after the observations from SEC-MALS.

Future Work

The objective is still to determine the 3D-structure of CDTb at high resolution, particularly the mature CDTb'' form. The current crystallisation conditions that have yielded diffraction quality crystals should be focused on to optimise crystal growth to obtain higher resolution diffraction from the crystals. High concentrations of pure protein are required for crystal growth, however SEC-MALS data and electron microscopy data have confirmed that CDTb'' is forming interchangeable oligomers of high molecular mass at high concentration. Therefore, efforts should be focused on stopping this sporadic oligomerisation during concentration, as this oligomerisation may prevent protein from forming tightly packed crystals, hence the lack of high resolution diffraction from CDTb''. In order to prevent this oligomerisation, a new range of buffers should be used with a variety of different detergents commonly used for membrane proteins.

Chapter 4

CDTa

4.1 CDTa-actin and CDTa-CDTb Complexes

4.1.1 Introduction

CDT

Clostridium difficile has a number of virulence factors that contribute to its pathogenicity, but this chapter will focus on one of the secreted toxins. The pathogenic strains of *C. difficile* produce two potent exotoxins: Toxin A and Toxin B (often called TcdA and TcdB) which both induce mucosal inflammation and diarrhoea. In addition to these exotoxins, some *C. difficile* strains produce an ADP-ribosylating binary toxin (CDT), however, the role of this toxin in disease is unclear.

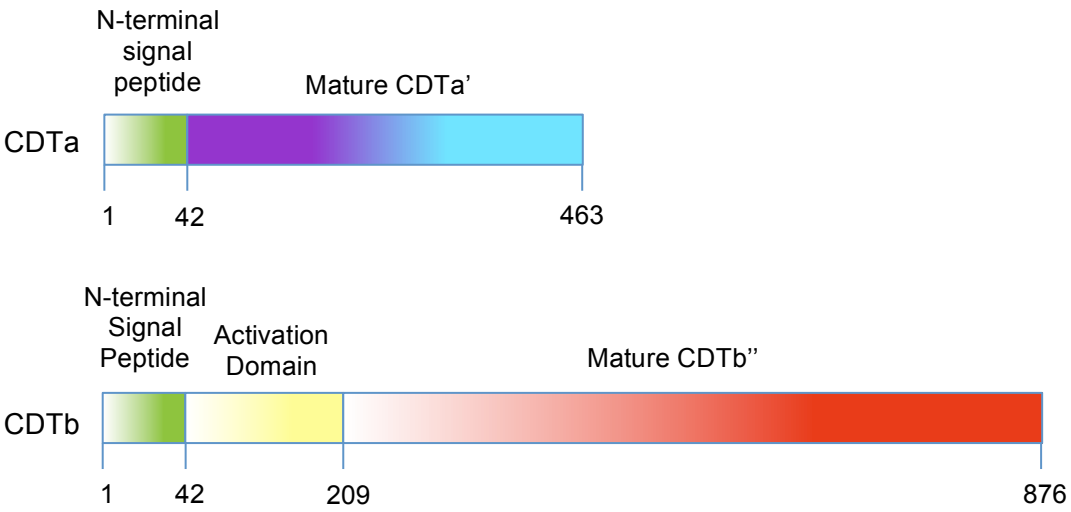


Figure 4.1.1a- Schematic representation of the domain organisation of the two components: CDTa and CDTb of the *C. difficile* ADPRT binary toxin. The N-terminal signal peptides are displayed in green for both components. The active CDTa' domain is coloured with the N-terminal domain (purple) and the C-terminal domain (cyan). The CDTb component has an activation domain (yellow) that must be cleaved to give the activated domain (red).

CDT is a member of the ADPRT family of binary toxins (Popoff *et al.*, 1988). It is a Type IV ADPRT meaning it is made up of two independently

produced components; the enzymatic component, CDTa (48kDa), and the transport component, CDTb (99 kDa). The enzymatic component irreversibly ADP-ribosylates G-actin, which reduces the ability of actin to polymerise into F-actin. This disruption of actin polymerisation leads to disturbance of the cytoskeletal architecture, which can lead to cell death. The closely related homologues of CDT: C2 toxin from *C. botulinum* and Ia toxin from *C. perfringens,* that are both Type IV ADPRTs have been shown to ADP-ribosylate G-actin at residue Arg-177 (Vandekerckhove *et al.*, 1988, 1987). Therefore we can predict that CDT behaves in a similar manner and will also ADP-ribosylate G-actin at Arg-177. There is poor understanding of this group of ADPRTs structurally as whole functioning toxins, and little understanding of the individual components

The focus of this chapter is to elucidate where possible more information about the enzymatic component, CDTa, in terms of both mechanism of ADP-ribosylation and interactions with CDTb and actin. For clarity in this chapter, the fully activated cytotoxic enzymatic component will be denoted CDTa (instead of CDTa'). The crystal structure of the CDTa component was solved previously at three different pH values, 4.0, 8.5 and 9.0 (PDB codes: 2WN8, 2WN4 and 2WN5 respectively). In addition the structure was solved in complex with ADP ribose donors NADPH and NAD at pH 9.0 (PDB codes: 2WN6 and 2WN7). The structure in complex with NAD was determined at 2.25Å and detailed mechanistic implications of CDTa have been proposed based on this structural data (Sundriyal *et al.*, 2009). The structure of CDTa in complex with NAD is displayed in Figure 1.2.2b.

The overall structure of CDTa is similar to that of its closest homologues; Ia from *C.perfringens* (PDB code: 1GIQ) and C2-I from *C.botulinum* (PDB code: 2J3X), as illustrated in Figure 4.1.1b (Schleberger *et al.*, 2006; Tsuge *et al.*, 2003). The core catalytic domain of the ADPRT structures is observed which consists of a network of perpendicular β-strands flanked by alpha helices. The active site structural motifs are located in equivalent positions in all three of the structures and are colour coordinated as demonstrated in Figure 1.2.2b. In order to compare the structural differences between these three homologues, the structural coordinates of Ia were superimposed onto those of CDTa (Figure 4.1.1c-A), and C2-I coordinates were separately superimposed onto CDTa (Figure 4.1.1c-B). As clearly demonstrated in Figure 4.1.1c, Ia and CDTa are structurally more similar than C2-I and CDTa.

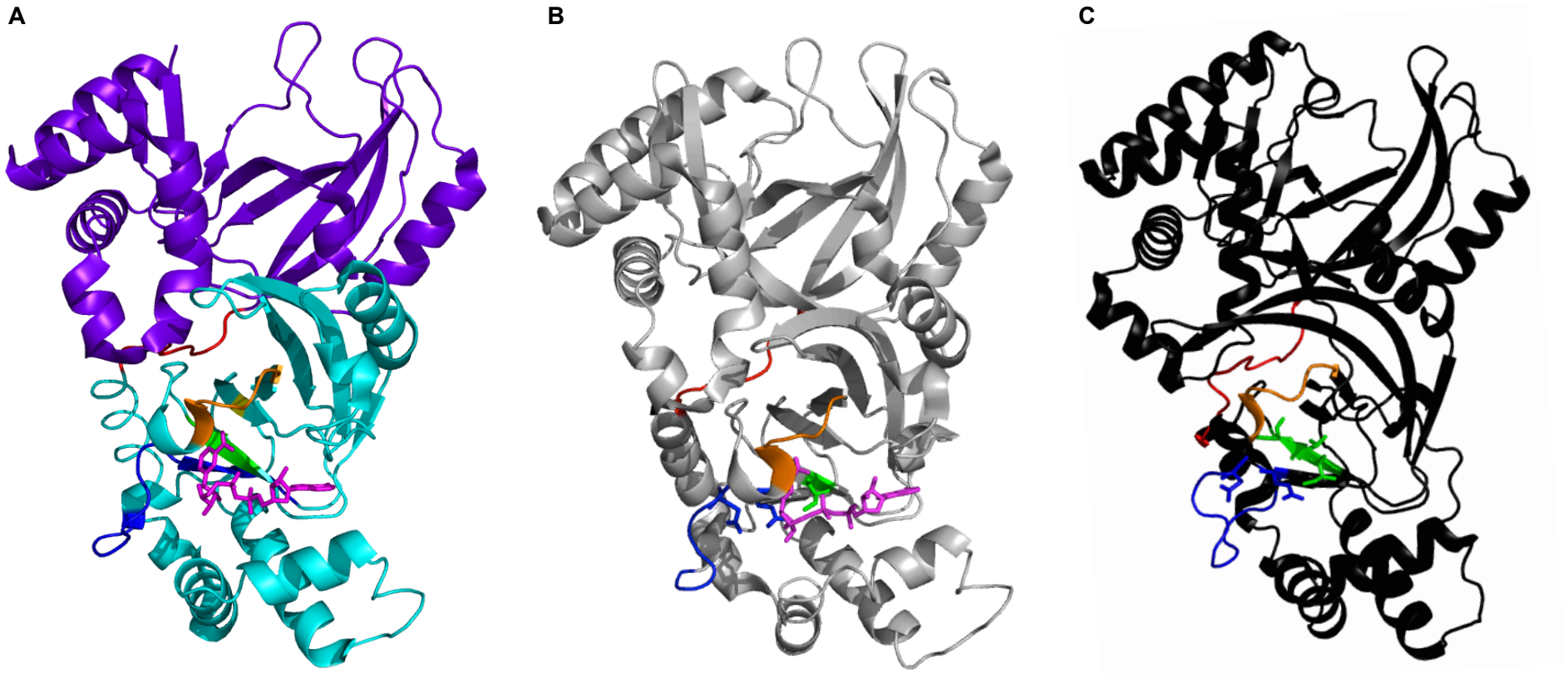


Figure 4.1.1b- CDTa, Ia and C2-I structures. A: The structure of CDTa with NAD bound (Sundriyal *et al.*, 2009)**, with the various structural motifs of the active site displayed in colour: substrate NAD in magenta, STS-motif in green, ARTT-loop and ExE-motif in blue, and the PN-loop in orange. The N-terminal domain is coloured purple, the C-terminal domain is coloured cyan and their connecting loop is coloured red. B: The structure of Ia with NAD bound** (Tsuge *et al.*, 2003)**, with the corresponding characteristic structural motifs coloured in accordance with CDTa. C: The structure of C2-I** (Schleberger *et al.*, 2006) **with the corresponding characteristic structural motifs coloured in accordance with CDTa. Images rendered using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

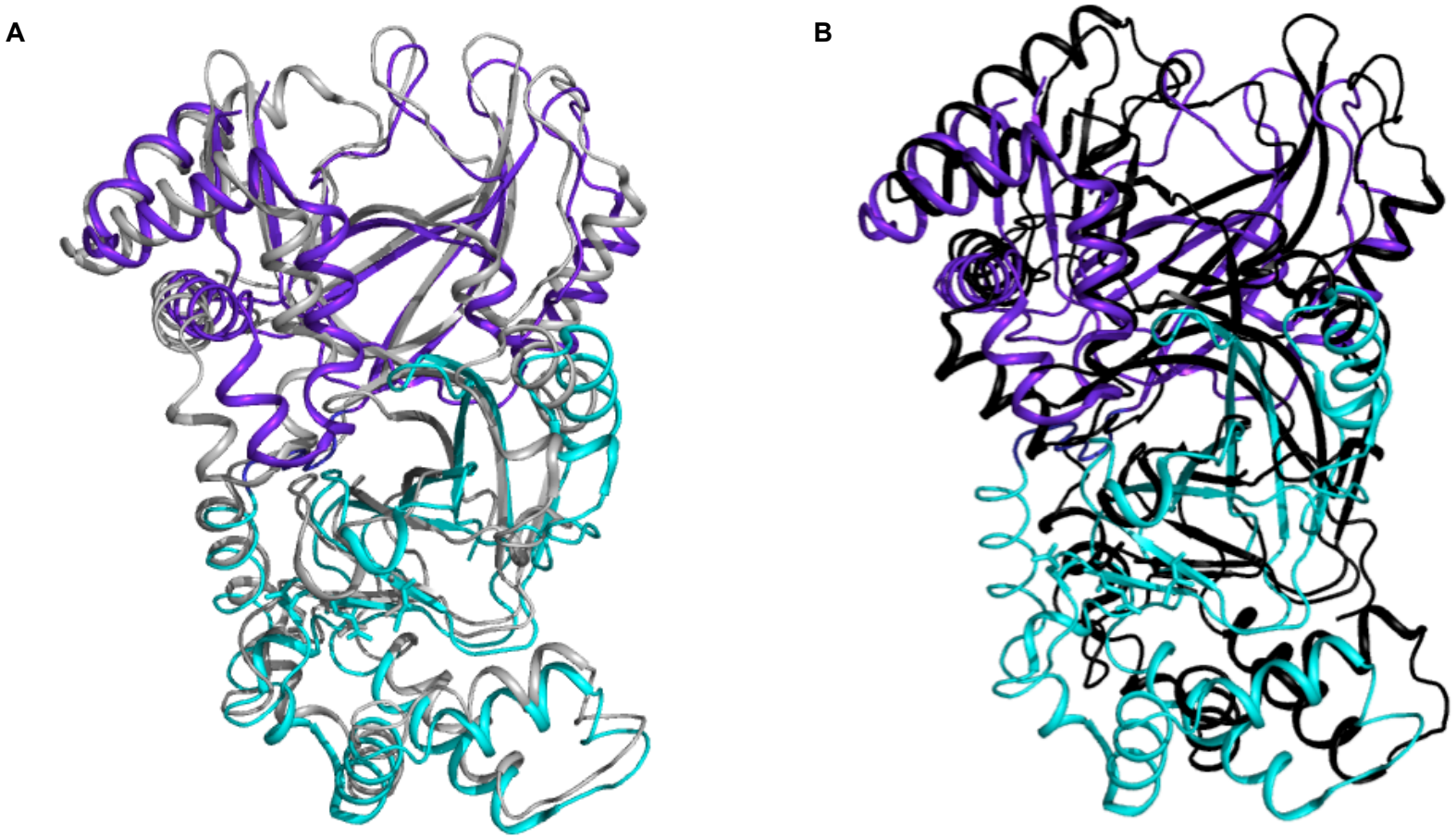


Figure 4.1.1c- Root mean square deviation (RMSD) of CDTa superimposed onto Ia (A) and C2-I (B). **A: Structure of CDTa superimposed onto the structure of Ia revealing only small variations in the structure (RMSD 0.84 Å) with 80 % sequence identity.** **B: Structure of CDTa superimposed onto the structure of C2-I revealing more structural deviations than that of the Ia toxin (RMSD 2.05 Å) with 40 % sequence identity. Images rendered using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

As CDTa and Ia are structurally more similar than C2-I, particularly in the catalytic core region, we can use the structural information known about Ia and the structure of Ia in complex with actin (see Figure 4.1.1d) to predict how CDTa will behave in terms of mechanism of ADP-ribosylation and the interactions with the target actin.

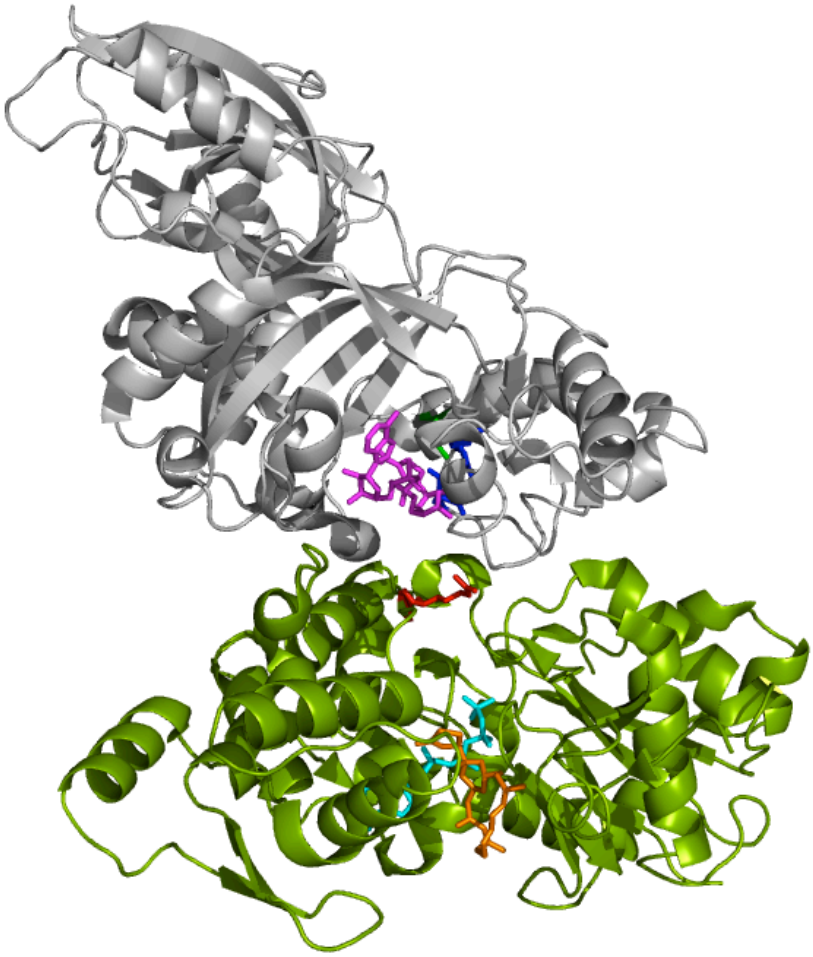


Figure 4.1.1d- The structure of the Ia-actin complex (Tsuge *et al.*, 2008)**. The Ia component is displayed in grey, with the non-hydrolysable NAD analogue (β-TAD) displayed as sticks in magenta. The ExE-motif and STS-motif of Ia are displayed as sticks in blue and green respectively. Actin is displayed in green with ATP bound in cyan and Latrunculin A displayed in orange. The Arg-177 residue of actin that is ribosylated during ADPRT activity is displayed as sticks in red. Images rendered using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

Mutagenesis studies performed for Ia revealed that the ExE-motif glutamic acid residues are essential for both NAD binding and hydrolysis (NADase activity) and ADP-ribosylation (ARTase) (Nagahama *et al.*, 2000). Therefore, we predict that the equivalent residues, Glu-385 and Glu-387 in CDTa will also be

essential for NADase and ARTase activity. We can therefore deduce the following mechanism for ADP-ribosylation of actin, see Figure 4.1.1e (Davies *et al.*, 2011):

A. NAD enters the active site where the ribose group is positioned in proximity to the catalytic Glu-387 residue.
B. The catalytic glutamate Glu-387 will form a H-bond with the 2'OH group of ribose and the nicotinamide group leaves.
C. An oxocarbenium intermediate is formed which is susceptible to nucleophilic attack. The ARTT-loop is rearranged to bring the Glu-385 residue into the reaction centre stabilising the transfer of ADP-ribose to Arg-177 on actin.
D. ADP-ribose is transferred to Arg-177.

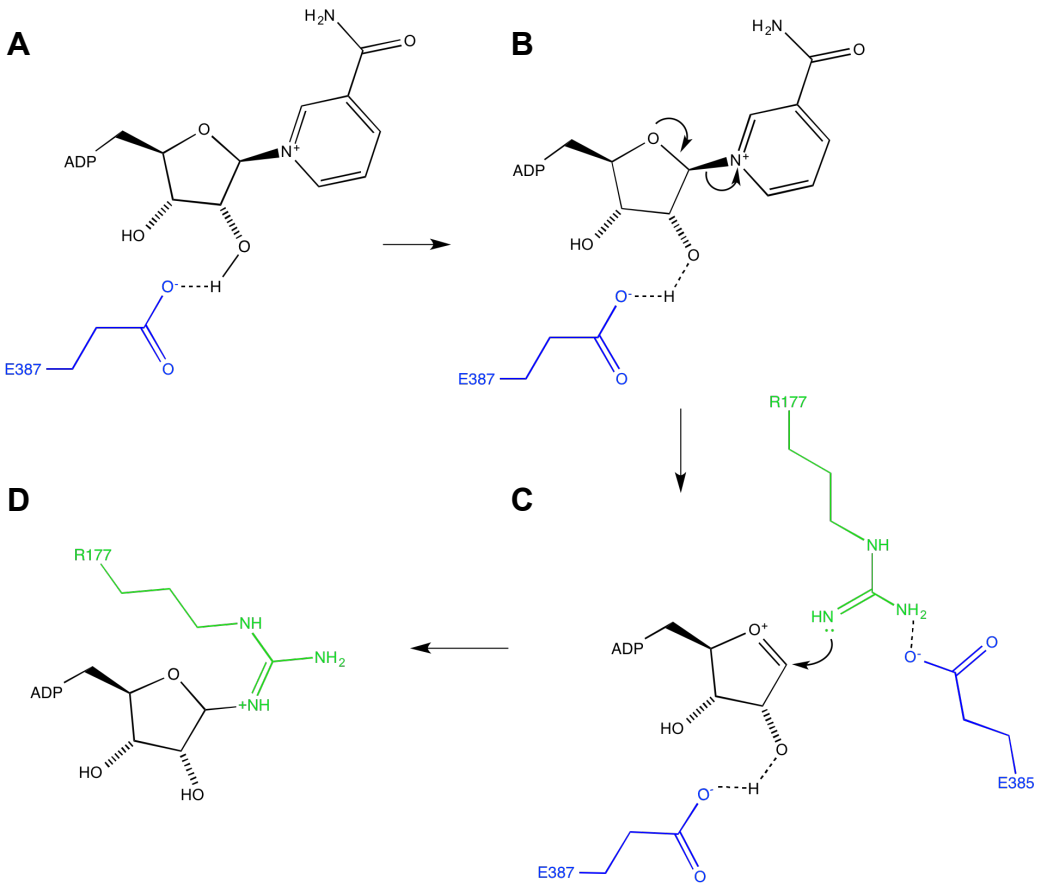


Figure 4.1.1e- Predicted mechanism for ADP-ribosylation of actin by CDTa. A: H-bond formation between E387 (blue) and 2 OH' of ribose (black). B: Loss of the nicotinamide group. C: E385 (blue) brings R177 (green) into active site and the ribose group oxocarbenium ion is susceptible to nucleophilic attack. D: R177 is ADP-ribosylated by CDTa. Figure adapted from Davies *et al.* (Davies *et al.*, 2011)**. Image creating using ChemDraw (2007 PerkinElmer Informatics).**

Further structural and experimental evidence is required to validate this hypothesis, which will be discussed in the following subchapters.

For CDTa the aims for this section were as follows:

4.1: Firstly to form a CDTa-actin complex and solve the structure and secondly to form a CDTa-CDTb complex and solve the structure.

4.2: To test the hypothesis of the mechanism by which CDTa ADP-ribosylates actin.

The above aims and outcomes will be discussed in subchapters 4.1 and 4.2 respectively. In order to have performed the experiments for the following chapters it was necessary to have a large quantity of pure CDTa, however the established protocols were limited in terms of yield. Therefore this present chapter will be focused on the optimisation of the established protocol for CDTa expression and purification, and the attempts made at complex formations.

4.1.2 Materials and Methods

Pre-existing Protocols

Expression of CDTa

The expression protocol for CDTa has previously been established (Sundriyal *et al.*, 2010). In brief, *Escherichia coli* BL21 (DE3) codon plus competent cells (Novagen ®) were transformed with pMAL-p2x-cdtA. Expression was carried out using Terrific broth (TB) medium supplemented with 1X TB salts, 0.5 % glucose and 100 µg/ml ampicillin. The TB medium was inoculated with 10 % overnight grown culture in Luria-Bertani (LB) medium supplemented with 100 µg/ml ampicillin. The cultures were allowed to grow at 37 °C until they reached an OD_{600}= 0.6-0.8 whereupon the temperature was then reduced to 20 °C. Protein expression was induced using 1 mM IPTG and cultures were harvested after 4 hours of expression, by centrifugation at 8,500 rpm for 10 min at 4 °C.

Purification of CDTa

The purification protocol for CDTa has previously been established (Sundriyal *et al.*, 2010). In brief, *E. coli* BL21 (DE3) codon plus cells containing expressed CDTa were lysed in buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl and 5 mM $CaCl_2$) using a French press in two cycles and cell debris was removed by centrifugation at 75,000 xg for 30 min at 4 °C. The supernatant was loaded onto Q-Sepharose Fast Flow resin (GE-Healthcare) pre-equilibrated with the lysis buffer. The resin was then washed until the absorbance at 280 nm base line was flat and stable. Protein was then eluted in 10 % of elution buffer (50 mM Tris-HCl, 1000 mM NaCl, and 5 mM $CaCl_2$) and sufficient Factor Xa (Novagen) was added to cleave the maltose binding protein (MBP) tag at 20 °C for 24 hour. The protein was then dialysed over night against 50 volumes of dialysis buffer (50 mM Tris-HCl pH 8.0 and 20 mM NaCl) at 20 °C. The protein was reloaded onto Q-Sepharose Fast Flow resin pre-equilibrated with dialysis buffer and the cleaved CDTa was collected in the flow through. CDTa was stored at -80 °C until further required.

Optimisation of Expression and Purification of CDTa

Initial Attempts at Optimisation of Purification

The above purification protocol resulted in low yield of pure CDTa. The following adjustments were implemented to optimise the purification protocol:

1. Use MBP column instead of Q-Sepharose resin, as the expressed CDTa has an MBP tag.
2. Try Source-Q resin, as it is designed to be more effective than Q-Sepharose resin.
3. Try using various additives such as 1 mM EDTA and 1 mM DTT.
4. Use alternative lysis technique such as bead beating instead of lysis by French press.
5. Use batch-binding technique instead of flow binding.
6. Clone CDTa DNA into a different vector and retry expression and purification tailored to the new plasmid.

Cloning of CDTa into Alternative Vector

CDTa DNA in vector pMAL-p2x was digested using BamH1 and Sal1. Concurrently pGEX-6p1 was digested with the same enzymes. The digest was analysed by an 0.8 % agarose gel, which was run at 120 V for 1 hour. The CDTa dropout bands were excised and gel purified (Wizard ® genomic DNA purification kit, Promega). The gel purified CDTa dropout DNA was then ligated into the digested pGex-6p1 vector using T4 DNA ligase. This was repeated twelve times to ensure ligation success of at least one reaction. The reactions were left at 22 °C for 30 min, and were then tested for the insert by performing rapid digests with BamH1 and Sal1.

Successfully ligated DNA was initially sequenced and then transformed into Top10 cells, using standard transformation techniques. Colonies of Top10 cells containing pGex-6p1-cdtA were used to inoculate an overnight culture in LB medium. The pGex-6p1-cdtA DNA was extracted from the overnight culture of

Top10 cells and was rapidly digested to check for the presence of the CDTa DNA insert, which was then analysed by an 0.8 % agarose gel. Where the dropout was present, the DNA was sent for sequencing to ensure successful ligation into the pGex-6p1 vector.

Optimised Expression of CDTa

E. coli BL21 (DE3) codon plus cells were transformed with the pGEX-6p1-cdtA expression construct. Overexpression of soluble CDTa was obtained using 2 L baffled flasks containing 1 L of LB medium supplemented with 100 µg/ml ampicillin. The medium was inoculated with 10 ml overnight culture of the expression host, grown at 37 °C in LB medium supplemented with 100 µg/ml ampicillin. Once the culture reached OD_{600} = 0.8-1.0, the temperature was reduced to 20 °C. After the temperature reached 20 °C, 1 mM IPTG was added to induce expression of CDTa. The cultures were harvested after 4 hours, using 500 ml centrifuge tubes in a 10:500 rotor, spinning at 8,500 rpm for 10 min at 4 °C. The cell pellets were re-suspended in 2 ml/g of pellet in lysis buffer; 50 mM Tris-HCl pH 8.0 and 20 mM NaCl. The re-suspended pellets were then rapid frozen in liquid nitrogen and stored at -80 °C until further use.

Optimised Purification of CDTa

A cell pellet, stored at -80 °C from the batches of expressed protein, was thawed and diluted in lysis buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl and 1 mM EDTA) until a total volume of 10 ml/g of pellet was reached. The cells were then passed through a homogenizer twice at a pressure of 20 kpsi, and temperature of 4 °C. Throughout the duration of the purification from this point onwards the protein was kept on ice. The cell lysate was centrifuged at 25,000 rpm for 35 min at 4 °C and the resultant supernatant was filtered through both a mini-sart GF-pre-filter (Sartorius) and 0.45 µm millex-HA filter (Millipore). The filtered supernatant was then loaded onto a 5 ml GSTrap-FF column (GE Healthcare) at 0.5 ml/min, pre-equilibrated with lysis buffer. The column was then washed in lysis buffer until the absorbance at 280 nm reading reached a constant stable base line. The bound protein was eluted at 0.5 ml/min in 20 mM reduced glutathione in lysis buffer. A sufficient amount (1 U per 100 µg of

protein) of PreScission protease (GE Healthcare) was added to ensure complete cleavage of the GST tag. The protein was dialyzed against 100 vol of dialysis buffer (50 mM Tris-HCl pH 8.0 and 20 mM NaCl), using 12-14 kDa cut off dialysis tubing (Medicell International Ltd) and was allowed to dialyze and cleave overnight at 4 °C. Dialyzed protein was then centrifuged at 4,000 g to remove any precipitated protein and was re-loaded back onto the GSTrap-FF column, pre-equilibrated in lysis buffer, at 0.5 ml/min, collecting the pure CDTa in the flow through. The column was then washed until a constant stable base line was reached and the bound GST tag was eluted using 20 mM reduced glutathione in lysis buffer. After washing the column in 3 CVs of water followed by 3 CVs of lysis buffer, the above step was repeated to ensure complete separation of the cleaved GST tag. The final pure CDTa was concentrated to 5.0 mg/ml using 10 kDa cut-off concentrators (Amicon Ultra-15, Millipore), rapid frozen in liquid nitrogen and stored at -80 °C.

Formation of CDTa-Actin Complex

Attempt 1- Following Ia-Actin Protocol

Initially, the protocol used for the Ia-actin complex was followed (Tsuge *et al.*, 2008). For the Ia-actin complex, 15 mg/ml of rabbit skeletal actin was mixed with 0.34 mM Latrunculin A (laγA) in buffer 2mM Tris pH 8.0, 0.2 mM ATP, 0.5 mM DTT, 0.2 mM $CaCl_2$ and 0.02% NaN_3. This was then mixed with 17-20 mg/ml of Ia to which 0.5 mM β-TAD was added, kindly provided by Dr Victor Marquez, Centre for Cancer Research, NCI-Frederick. LayA is a small molecule that prevents actin from polymerising and β-TAD is a non-hydrolysable NAD analogue that helps stabilise the complex. As we have a much lower yield of CDTa, the same ratio of components were combined but at lower concentrations. The protocol was performed in an identical manner, 31 nM was combined with 29 nM laγA in buffer 2mM Tris pH 8.0, 0.2 mM ATP, 0.5 mM DTT, 0.2 mM $CaCl_2$ and 0.02% NaN_3 which was then combined with 36 nM pure CDTa and 43 nM β-TAD. As with the Ia-actin compex, no analytical steps were taken to ensure complex formation, however, the components were incubated overnight at 4 °C. The following five crystallisation screens were performed: Morpheus HT-96,

JCSG-plus HT-96, PACT premier HT-96, ProPlex HT-96 and Structure screen I + II HT-96.

Upon observation of any crystal hits, crystals were taken to Diamond Light Source, mounted in LithoLoops (Hampton Research) and one at a time, the LithoLoops were mounted on a goniometer and were rotated whilst being exposed to a beam of X-rays.

Complex	Enzyme	Actin	Latrunculin A	β-TAD
Ia-actin	420 nM	360 nM	340 nM	500 nM
CDTa-actin attempt 1	36 nM	31 nM	29 nM	43 nM
CDTa-actin attempt 2	36 nM	31 nM	29 nM	43 nM
CDTa-actin attempt 3	49 nM	64 nM	64 nM	64 nM
CDTa-actin attempt 4a	15 μM	23 μM	30 μM	30 μM
CDTa-actin attempt 4b	15 μM	18 μM	100 μM	48 μM

Table 4.1.2a- Different ratios of complex components used during attempts at forming a CDTa-actin complex.

Attempt 2- Adding a Gel-Filtration Step

For the second attempt at CDTa-actin complex formation, the same ratio of components were added in a similar manner, see Table 4.1.2a, however the mixture was incubated overnight at 16 °C instead of 4 °C, as this is the temperature at which crystallisation plates would be set up. As a precaution, the complex mixture was passed through a superdex-200 10/300 GL column (GE Healthcare) to check for the presence of the complex. As CDTa and actin have

similar molecular weights, we would expect to see a single peak from a gel-filtration column if no complex had formed, and potentially two peaks, one slightly shifted left than the other, if there was complex formation. The complex mixture was concentrated to 1 ml and was loaded onto a Superdex-200 10/300 gel-filtration column using a 2 ml loop. The flow rate was set to 0.2 ml/min and 0.5 ml fractions were collected. Analysis of complex formation by SDS-PAGE also proved difficult due to the similarity in molecular weights of CDTa and actin, and also because the complex would break apart during sample preparation for the gel.

Attempt 3- Alternative Ratio of Components and Method of Combining

For the third attempt at CDTa-actin complex formation an alternative ratio of components was trialled, see Table 4.1.2a. In this attempt, the β-TAD was directly added to CDTa and was incubated at room temperature for 30 min. Actin was dissolved in the complex buffer: 2mM Tris pH 8.0, 0.2 mM ATP, 0.5 mM DTT, 0.2 mM $CaCl_2$ and 0.02% NaN_3, and IaγA was immediately added to prevent polymerisation. Actin and IaγA were incubated at room temperature for 30 min. After the incubation period, all of the components were mixed together and were incubated overnight at 4 °C, with gentle shaking. The sample was concentrated down to 1 ml and loaded onto a Superdex-200 10/300 GL column (GE Healthcare) using a 2 ml loop. The flow rate was set to 0.2 ml/min and 0.5 ml fractions were collected. The fractions were analysed by SDS-PAGE.

Attempt 4a- Try Complex Formation on Column

As the previous three attempts have been unsuccessful, an alternative method of complex formation has been implemented. The CDTa-GST protein was taken after the first step of the optimised purification protocol and was applied to GST-resin using batch binding: CDTa-GST was mixed with unpacked resin and incubated overnight at 4 °C, with gentle shaking. The following day the mixture was applied to a gravity column and the any unbound protein was washed off the resin using complex buffer (as above). The following components

were mixed together; 5 mg actin, 30 nM β-TAD and 30 nM IaγA in complex buffer and were then applied to the GST-resin containing bound CDTa-GST, by resuspension of the resin with gentle rocking. The mixture was incubated for 1 hour at 4 °C, with gentle shaking. As the GST-tag is fused at the N-terminus, there should be no effect on actin binding. The resin containing the complex mixture was then reapplied to a gravity column and any actin that did not bind CDTa was washed off using complex buffer. The resin was then resuspended in complex buffer, and sufficient amount of PreScission Protease (70 µl of 2U/µl stock) was added to the resin to remove the GST-tag. The cleaved mixture was reapplied to a gravity column and any CDTa-actin complex was collected in the flow through. The bound GST-tag was eluted using complex buffer containing 20 mM reduced glutathione. Samples were taken throughout the procedure and were analysed by SDS-PAGE, see Figure 4.1.3i.

Attempt 4b- Try Complex Formation on Column

Attempt 4a was repeated with some minor alterations: a slightly different ratio was used, see Table 4.1.2a and the incubation period with all components was extended to overnight at 4 °C rather than 1 hour. Samples were taken throughout the procedure and were analysed by SDS-PAGE, see Figure 4.1.3j.

Formation of a CDTa-CDTb Complex

Attempt 1

CDTa was purified using the optimised protocol from this chapter and CDTb'' was purified using the optimised protocol from Chapter 3. The first attempt made at formation of CDTa-CDTb complex was performed using a simple method of combining the two proteins in a 1:1.2 (CDTa:CDTb'') molar ratio, incubating overnight at 4 °C, then performing a gel-filtration step. The details of the experiment are as follows: 15 µM of CDTa was combined with 18 µM CDTb'' in a total reaction volume of 250 µl. The mixture was left to incubate overnight at 4 °C with gentle rocking. A superdex-200 10/300 (GE Healthcare)

was pre-equilibrated with complex buffer: 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 2 mM DTT and 10 % glycerol. The sample was loaded onto the superdex-200 10/300 column using a 1 ml loop and was left to run at 0.2 ml/min collecting 500 µl fractions. The fractions were analysed by SDS-PAGE, Figure 4.1.3k.

Attempts 2 and 3

In the second attempt, the same protocol was followed except alternative ratios of the two components were trialled: 1:3.5 molar ratio of CDTa:CDTb'' (specifically 13 µM CDTa + 46 µM CDTb''). In addition, 45 mM beta-octyl-glucoside (β-OG) was added with the intention of stabilising the CDTb'' oligomer with this detergent. In this second attempt, the components were incubated for 1 hour at 4 °C prior to the gel-filtration step. In the third attempt, a 1:1 molar ratio was trialled (specifically 23 µM + 23 µM) in the presence of 45 mM β-OG again, but was left to incubate overnight at 4 °C.

4.1.3 Results

Expression and Purification of CDTa Using Established Protocols

Expression of CDTa from pMAL-p2x-cdtA

The previously established protocol (Sundriyal *et al.*, 2010) for the expression of CDTa was used to obtain enough protein for later use in the formation of a CDTa-actin complex. *E.coli* BL21 (DE3) codon plus cells were transformed with pMAL-p2x-cdtA, which were then grown in LB medium supplemented with ampicillin, at 37 °C until they reached an OD_{600} = 0.6-0.8. Expression was induced using 1 mM IPTG at 20 °C. Cells were then harvested after 4 hours. Pre-induction and post-induction samples were taken for analysis by 10 % Tris-Glycine SDS-PAGE.

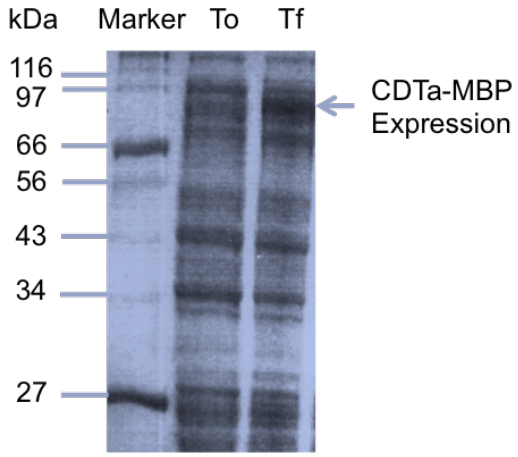


Figure 4.1.3a- Expression of CDTa-MBP in *E.coli* BL21 (DE3) codon plus cells. Lane 1 on the 10 % Tris-Glycine SDS-PAGE displays the pre-stained ladder, lane 2 displays the pre-induced cell sample and lane 3 displays the post-induction expression sample. The recombinantly expressed soluble CDTa-MBP can be seen as indicated by the arrow at approximately 86 kDa.

As seen in Figure 4.1.3a, the post-induction sample shows good levels of expression of CDTa-MBP in comparison to the pre-induction sample. Therefore, purification of CDTa-MBP was performed.

Purification of CDTa Using Anion Exchange Chromatography

The previously established purification protocol (Sundriyal *et al.*, 2010) was followed to obtain a high quality of purity for use in formation of a CDTa-actin complex. In brief, BL21 (DE3) codon plus cells containing CDTa-MBP were lysed using a French press and centrifuged to remove remaining cell debris. The cell lysate was loaded onto Q-sepharose resin and bound CDTa-MBP was eluted using increased salt concentration. The MBP-tag was cleaved using Factor Xa, and the protein was dialysed against a large volume of buffer to remove the salt. The cleaved protein was then loaded back onto Q-sepharose resin, to which the MBP-tag binds and the pure CDTa was collected in the flow through.

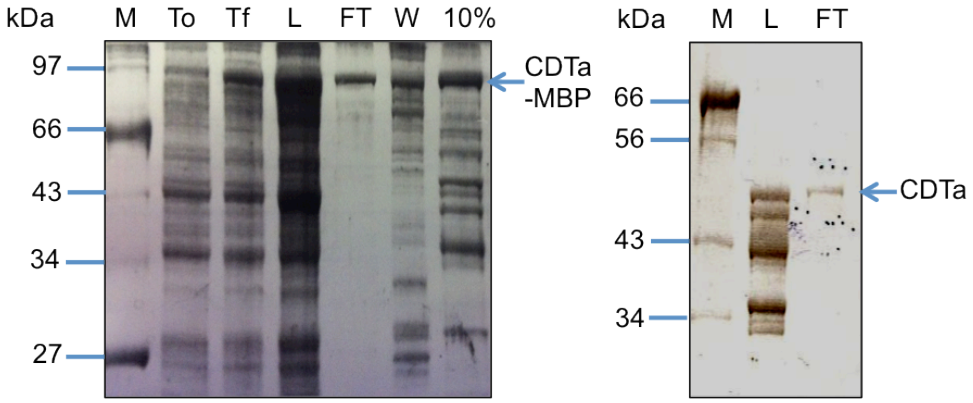


Figure 4.1.3b- 10 % Tris-Glycine SDS-PAGE displaying the steps taken in the purification of CDTa. The first gel (left) displays the pre-stained ladder in the first lane (M), the pre-induced expression sample in the second lane (To) and the expressed CDTa-MBP in the third lane (Tf). The fourth and fifth lane displays the loaded cell supernatant (L) and the non-bound protein flow through (FT), respectively. The sixth lane displays the wash step (W), whereby the resin is washed until the absorbance 280 nm reading reached a steady flat base line. Finally the last lane on the first gel (left) shows the elution step in 10 % elution buffer (10%). The second gel (right) displays the pre-stained protein ladder in the first lane (M), the cleaved protein loaded onto the resin (L) and the pure CDTa in the third lane (FT) indicated by the arrow.

As displayed in Figure 4.1.3b, the purification protocol works effectively to produce pure CDTa. However, the CDTa yield is poor, with only 0.2 mg of CDTa obtained from a 1 L culture. Therefore, a number of adjustments were made to attempt to improve the final yield of pure CDTa.

Optimisation of Expression and Purification of CDTa

Expression of CDTa from pGEX-6p1-cdtA

The following table displays the alterations that were made to the expression and purification protocols in attempts to optimise the production of CDTa, with the corresponding results.

Modification to Protocol	Result
MBPTrap column used instead of Q-sepharose resin	CDTa-MBP did not bind the MBPTrap column
Source-Q resin used instead of Q-sepharose resin	No significant improvement in yield
1 mM EDTA and 1 mM DTT additives	No effect on yield
Bead beating cell lysis instead of French press	Some improvement in yield, but not significant
Batch-binding to Q-sepharose resin instead of flowing through pre-packed column	Some improvement in yield, but not significant
Clone DNA into alternative expression vector	Significant improvement in quality and quantity of CDTa yield, detailed below

Table 4.1.3a- Optimisation of CDTa expression and purification. The table displays modifications made to protocols in attempts to improve CDTa yield.

The final details of the optimised CDTa expression and purification protocols will be given below.

Cloning of CDTa DNA into Alternative Expression Vector

The plasmid pMAL-p2x-cdtA was digested using BamH1 and Sal1 restriction enzymes, and was then ligated into the new expression vector pGEX-6p1, which had also been digested with the same restriction enzymes. The

ligation was repeated twelve times, to ensure success of at least one ligation. After ligation, each reaction mixture was digested again to check for the insert prior to transformation, see Figure 4.1.3c for results.

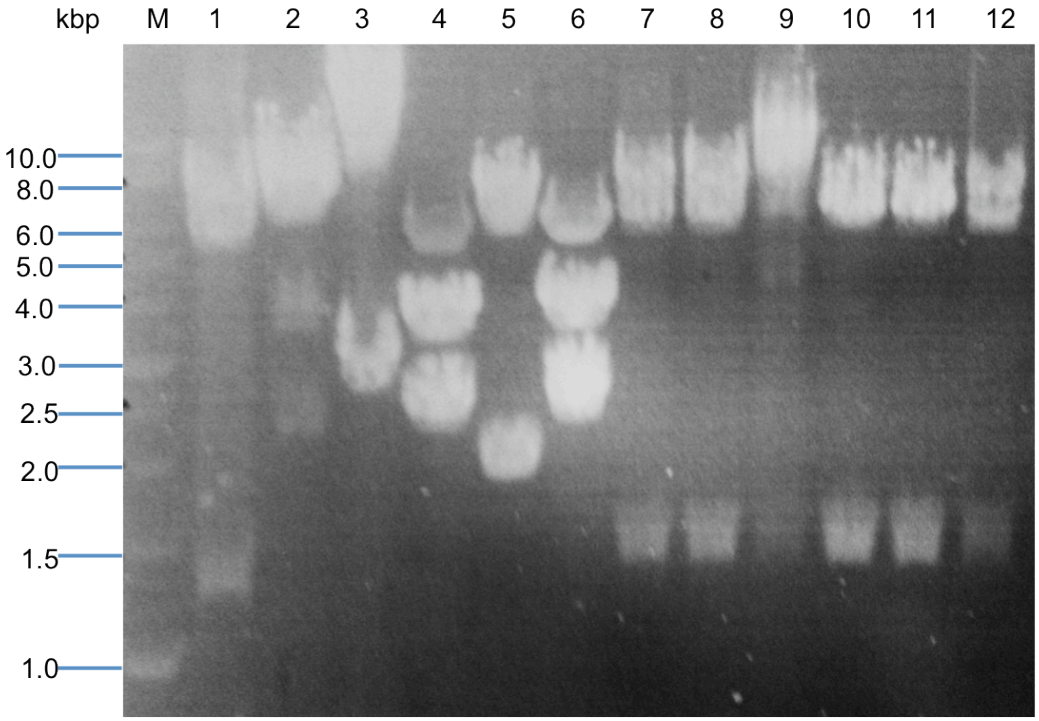


Figure 4.1.3c- 0.8 % agarose gel displaying the quick digest for ligation results. Lanes 1, 7, 8, 10-12 show the dropout CDTa DNA at approx. 1.5 kbp and the vector at approx 5 kbp, indicating successful cloning of CDTa DNA into pGEX-6p1.

The successfully ligated DNA was transformed into Top10 cells, cultured and the DNA was extracted using the miniprep kit (Wizard ® plus Minipreps DNA Purification System, Promega). After sequencing to confirm the CDTa DNA was in frame, BL21 (DE3) codon plus cells were transformed ready for expression.

Optimised Expression of CDTa Using a New Expression Vector

The CDTa DNA was cloned into an alternative expression vector, pGEX-6p1 and a new optimised expression protocol was established. In brief, *E. coli* BL21 (DE3) codon plus cells were transformed with the pGEX-6p1-cdtA expression construct and soluble CDTa expression was obtained using LB

medium supplemented with 100 µg/ml ampicillin. The expression host was grown at 37 °C until the culture reached OD_{600} = 0.8-1.0, the temperature was reduced to 20 °C and expression of CDTa was induced using IPTG. The cultures were harvested after 4 hours by centrifugation at 8,500 rpm for 10 min at 4 °C and cell pellets were re-suspended in lysis buffer, followed by rapid freezing in liquid nitrogen.

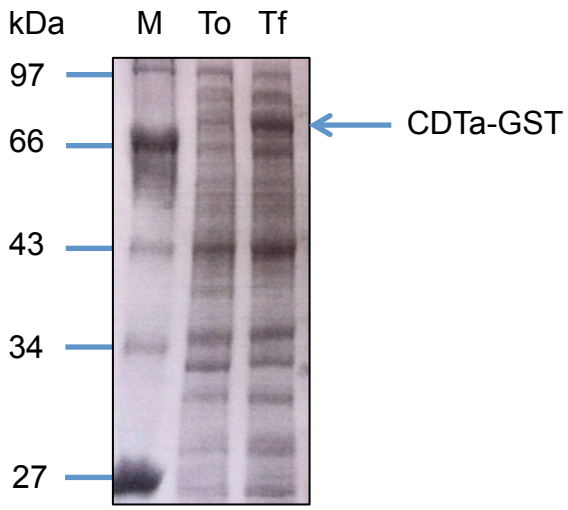


Figure 4.1.3d- 10 % Tris-glycine SDS-PAGE displaying expressed CDTa-GST. Lane 1 displays the prestained protein ladder (M), lane 2 displays the pre-induced culture sample (To) and lane 3 displays the post-induction sample (Tf) with CDTa-GST at 77 kDa.

As demonstrated in Figure 4.1.3d, the expression protocol is successful as a clear protein band can be observed at 77 kDa in the post-induced sample (Tf), that is not present in the pre-induced sample (To). Due to the expression success, the cell pellets containing CDTa-GST were lysed ready for purification of CDTa.

Optimised Purification of CDTa Using New Method

As CDTa was expressed with a GST-tag instead of an MBP-tag, alternative methods for purification were used. GST-affinity chromatography was

performed using a 5 ml GSTrap-FF column at low flow rate to ensure complete binding of the CDTa-GST to the resin.

In brief, cell pellets stored at -80 °C from the batches of expressed protein, were thawed and diluted in lysis buffer (50 mM Tris-HCl pH 8.0, 20 mM NaCl and 1 mM EDTA) until a total volume of 10 ml/g of pellet was reached. The cell lysate was then lysed and centrifuged to remove cell debris, and the resultant supernatant was filtered prior to loading the GSTrap-FF column pre-equilibrated with lysis buffer. The column was then washed and bound protein was eluted using reduced glutathione in lysis buffer. PreScission protease was added to ensure complete cleavage of the GST tag and the protein was dialyzed to remove the reduced glutathione. Cleaved protein was then re-loaded back onto the GSTrap-FF column, pre-equilibrated in lysis buffer, collecting the pure CDTa in the flow through. The bound GST tag was eluted using reduced glutathione in lysis buffer and after washing the column and pre-equilibrating the column again, the above step was repeated to ensure complete separation of the cleaved GST tag, see Figure 4.1.3e for results.

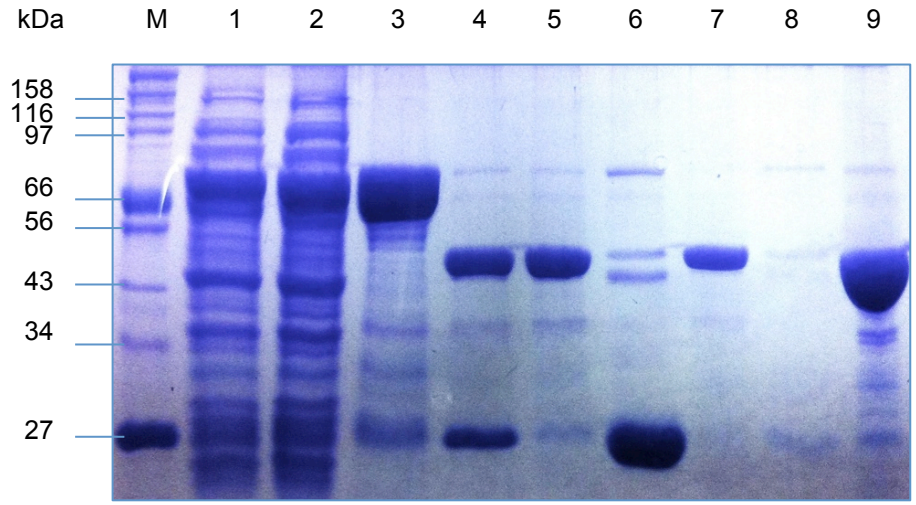


Figure 4.1.3e- Purification steps of CDTa displayed using 10 % Tris-Glycine SDS-PAGE. Lanes 1 and 2 represent cell lysate load and flow through respectively. Lanes 3 and 4 represent CDTa-GST elution from column and the overnight cleavage of GST-tag using PreScission Protease. Lanes 5 and 6 represent the 'pure' CDTa flow through and GST-tag elution respectively. Lanes 7 and 8 represent the repetition of the previous step to ensure complete removal of GST. Lane 9 represents the final concentrated CDTa.

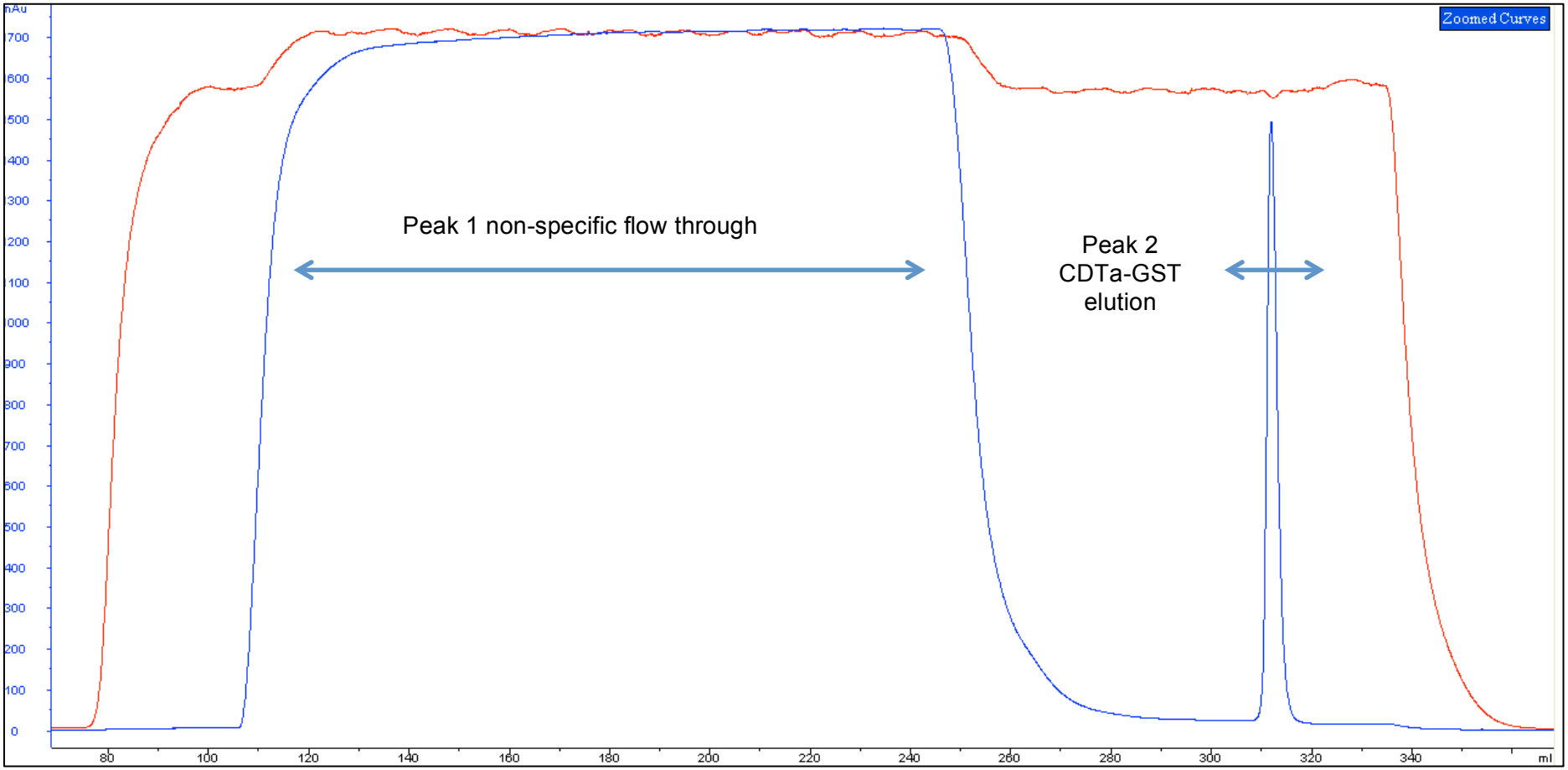


Figure 4.1.3f- Chromatogram for step 1 GST-affinity chromatography. **The first broad peak (Peak 1) represents non-specific flow through (>1600 mAu), and the second sharp peak (Peak 2) represents the eluted CDTa-GST protein (1500 mAu) using reduced glutathione elution. The blue line represents the absorbance at 280nm and the red line represents conductivity.**

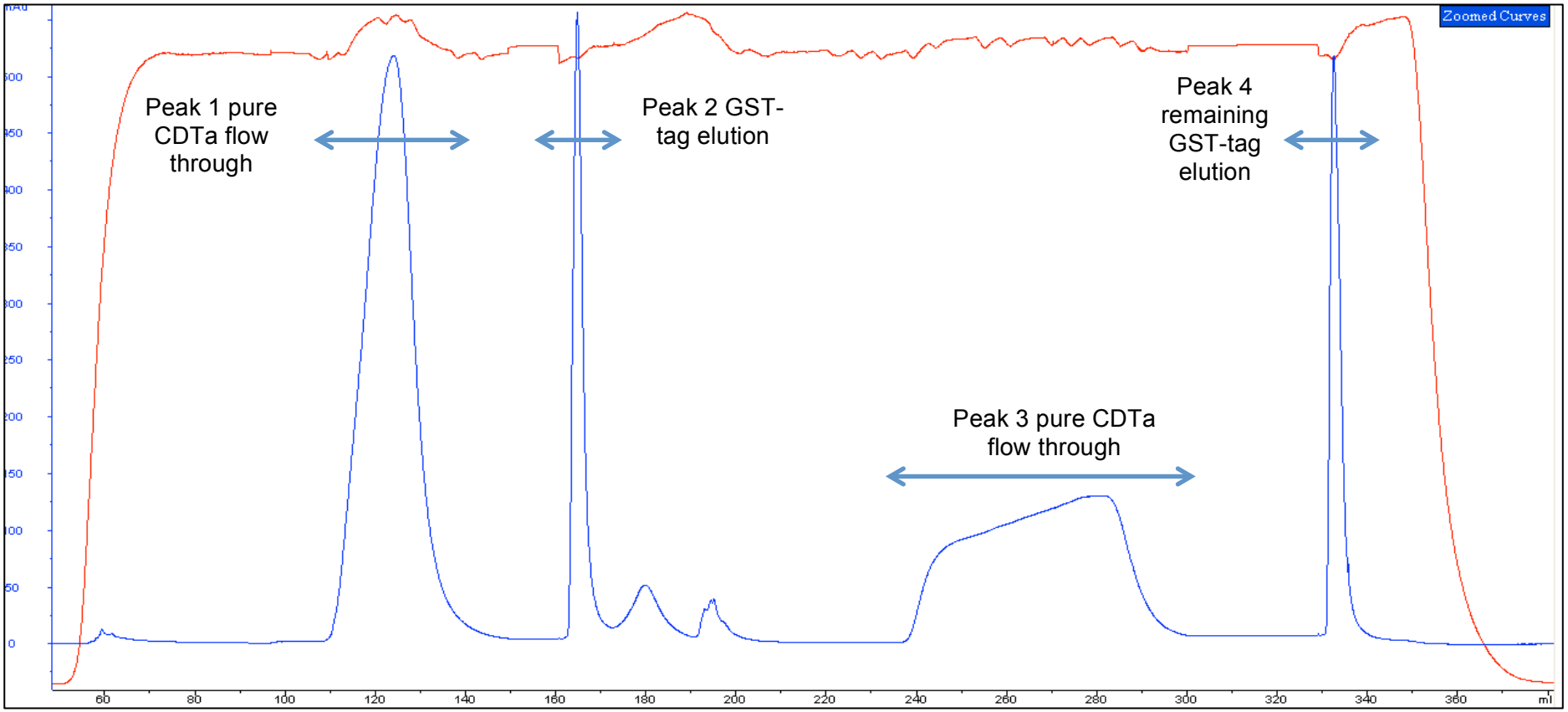


Figure 4.1.3g- Chromatogram for step 2 GST-affinity chromatography. **Peak 1 represents the CDTa that has had the GST-tag removed, which does not bind to the column. Peak 2 represents GST-tag that has bound to the column and has been eluted using 20 mM reduced glutathione. Peak 3 represents the pure CDTa that has been passed back through the column to ensure complete removal of the GST-tag. Peak 4 represents the remaining GST-tag separated from CDTa. . The blue line represents the absorbance at 280nm and the red line represents conductivity.**

The chromatogram in Figure 4.1.3f is an example of the first step from the CDTa-GST purification using a GSTrap-FF column. The first broad peak represents the flow through of proteins that have no affinity for the GSTrap resin. The column is then washed in buffer and the baseline is reached, at this point all of the non-specific proteins will have been washed off the column leaving only CDTa-GST bound to the column. The second peak represents the CDTa-GST elution peak, whereby reduced glutathione was applied to the column, which competes for GSTrap binding, allowing CDTa-GST to be eluted. The chromatogram in Figure 4.1.3g represents the second stage in the purification protocol whereby CDTa with GST-tag removed is passed back through the GSTrap-FF column to remove the GST-tag.

Formation of CDTa-actin Complex

Attempt 1

For the first attempt at forming a CDTa-actin complex, the protocol established for the Ia-actin complex was followed (Tsuge *et al.*, 2008). In brief: actin was mixed with IaγA, and both were then mixed CDTa and β-TAD. After an overnight incubation at 4 °C, the following five crystallisation screens were performed: Morpheus HT-96, JCSG-plus HT-96, PACT premier HT-96, ProPlex HT-96 and Structure screen I + II HT-96. In addition, a control crystallisation plate was set up using 5 mg/ml CDTa and the crystallisation condition that was used to grow crystals for which data were collected and the structure was solved (Sundriyal *et al.*, 2009). A number of crystallisation hits were observed which had different morphologies to the standard CDTa crystal morphology, see Figure 4.1.3h. The crystals were taken to Diamond Light Source, mounted in LithoLoops (Hampton Research) and one at a time, the LithoLoops were mounted on a goniometer and were rotated whilst being exposed to a beam of X-rays. Needle-like crystals of CDTa, Figure 4.1.3h-A, diffract to 2 Å resolution and have the same cell dimensions as the previously reported CDTa structure (Sundriyal *et al.*, 2009), see Table 4.1.3b. Hexagonal plate-like crystals of a possible CDTa-actin complex, displayed in Figure 4.1.3h-B also diffracted to high resolution, but show similar cell dimensions to that of the native CDTa, which suggests that complex formation has failed.

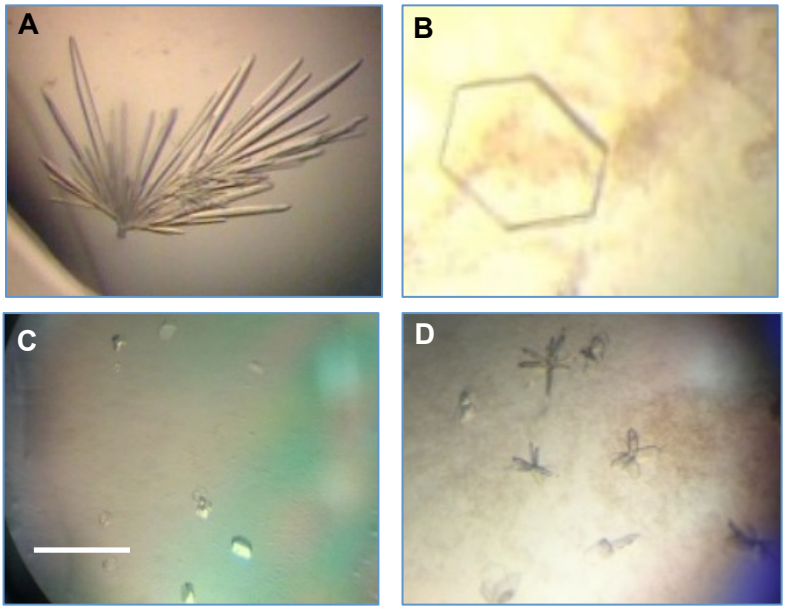


Figure 4.1.3h- CDTa crystals and possible CDTa-actin crystals. A: Standard CDTa crystals with needle-like morphology, grown in crystallisation condition 0.1 M MIB pH 9.0 and 20% PEG. B: Possible CDTa-actin hexagonal plate-like crystals, grown in condition 0.1 M Morpheus Buffer 1 pH 6.5, 0.12 M Ethylene glycols and 20 % PEG 550 MME and PEG 20k. C: Possible CDTa-actin small fragment crystals, grown in condition 0.1 M Morpheus Buffer 1 pH 6.5, 0.12 M Ethylene glycols and 30 % PEG 550 MME and PEG 20k. D: Possible CDTa-actin crystal clusters, grown in crystallisation condition 0.1 M Morpheus Buffer 1 pH 6.5, 0.12 M Ethylene glycols and 15 % PEG 550 MME and PEG 20k. The scale bar in image C represents 200 µm.

CDTa and CDTa-actin cell dimensions and space groups							
	a	**b**	**c**	**α**	**β**	**γ**	**Space Group**
CDTa crystal A	62.2	46.7	77.6	90.0	97.8	90.0	$P2_1$ (monoclinic)
CDTa-actin crystal B	57.4	45.0	77.3	90.0	101.8	90.0	$P2_1$

Table 4.1.3b- Cell dimensions calculated from the diffraction images collected during test shots for crystals A and B from Figure 4.2.3h. Both sets of cell dimensions are very similar suggesting complex formation has not been achieved.

Attempt 2

For the second attempt at CDTa-actin complex formation, the same ratio of components were added in a similar manner, however the mixture was incubated overnight at 16 °C instead of 4 °C, as this is the temperature at which crystallisation plates would be set up. A gel-filtration step was added to determine whether or not a complex had formed. This attempt was also unsuccessful at full complex formation, as only a large single peak was observed on the gel-filtration chromatogram at the predicted size of 40 kDa, which contained a mixture of CDTa and monomeric actin.

Attempt 3

For the third attempt at complex formation, an alternative ratio of components was trialled; see Table 4.1.2a. In order to saturate CDTa with β-TAD and actin, a high concentration of both components were added. Again, the mixture of components was concentrated and was passed through a gel filtration column to determine whether or not a complex had been formed. A shift in the peak towards the y-axis should be noted on complex formation, but unfortunately the same pattern was observed as with the previous attempt: a single sharp peak at approximately 40 kDa.

Attempt 4a

As the previous three attempts were unsuccessful, an alternative method of complex formation was implemented whereby attempts were made at forming an 'on column' complex. CDTa with a GST-tag was applied to a GSTrap column and was washed to ensure CDTa purity bound to the column. The GSTrap resin was re-suspended in a solution containing the remaining complex components and were incubated together overnight. As the GST-tag is fused at the N-terminus, there should be no effect of complex formation as actin is proposed to interact with the C-terminal domain. The resin was washed to ensure that anything bound to the resin was either CDTa-GST alone or a complex of CDTa-GST with actin. The components were then treated with PreScission Protease to cleave off the GST-tag allowing release of the potential complex from the resin and any excess CDTa. The reaction was then passed back through a GSTrap

column to remove any excess GST-tag. The results were analysed by 10 % Tris-glycine SDS-PAGE and 4-16 % Bis-Tris Native gel, see Figure 4.1.3i.

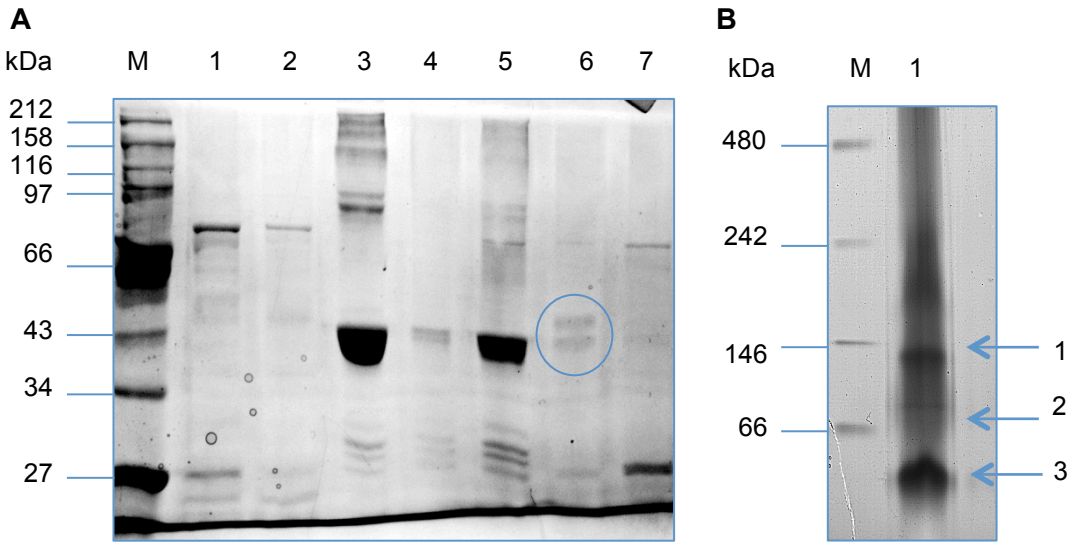


Figure 4.1.3i- Attempt 4a of CDTa-actin Complex Formation. A: The steps taken towards complex formation; lanes 1 and 2 represent CDTa-GST binding to the GST-resin, lane 3 represents the concoction of actin, IaγA and β-TAD, and lane 4 is the same concoction diluted by GST resin. Lane 5 represents the non-bound protein wash and lane 6 contains the PreScission Protease cleavage and therefore elution of any potential complex, and finally lane 7 is the remaining GST-tag elution from the column. B: 4-16 % Native Bis-Tris gel picture displaying lane 6 from gel A. Three clear bands of approximately 120 kDa (band 1), 80 kDa (band 2) and 40 kDa (band 3) are apparent on the gel.

Figure 4.1.3i displays the 'on column' batch binding method used to form the CDTa-actin complex. Up until now there has been no success in complex formation, however, these experiments show indications of small quantities of complex that are highly promising for future work. The 10 % Tris-glycine SDS-PAGE gel picture displayed in Figure 4.1.3i-A shows the steps taken to form a CDTa-actin complex, starting with CDTa-GST binding to the column, with the protein band apparent at approximately 77 kDa (lanes 1 and 2). Lane 3 indicates the concoction of excess actin containing IaγA and β-TAD and once it was applied to the resin containing CDTa-GST bound is indicated by lane 4. After the resin was washed to remove any unbound protein, a large quantity of actin was observed, see lane 5. However, surprisingly once the PreScission Protease was added to cleave off the GST-tag, equal volumes of proteins of approximately 45

kDa and 43 kDa were eluted from the resin. As they are in a ratio of approximately 1:1 this is a strong indicated that a CDTa-actin complex was formed. They may appear on this gel picture as two separate bands rather than one higher band because it is a reducing gel. To confirm the presence of the complex, the sample was analysed using a 4-16 % native gel, for which the results are displayed in Figure 4.1.2i-B. The key thing to note from this gel picture is the presence of a strong band of approximately 40 kDa (band 3) which could represent any excess CDTa or actin that have not formed a complex. Secondly there is a band of approximately 80 kDa (band 2), which is the correct size for a CDTa-actin complex, which is highly promising. Finally there is a band of approximately 120 kDa, which is not present on the 10 % reduced SDS-PAGE gel picture, but is present on the native gel (band 1). This size could mean we have a trimer of either CDTa or of actin.

Attempt 4b

In a second attempt using the 'on column' batch binding method, it was decided that in order to obtain a better yield of complex, a number of factors should be altered.

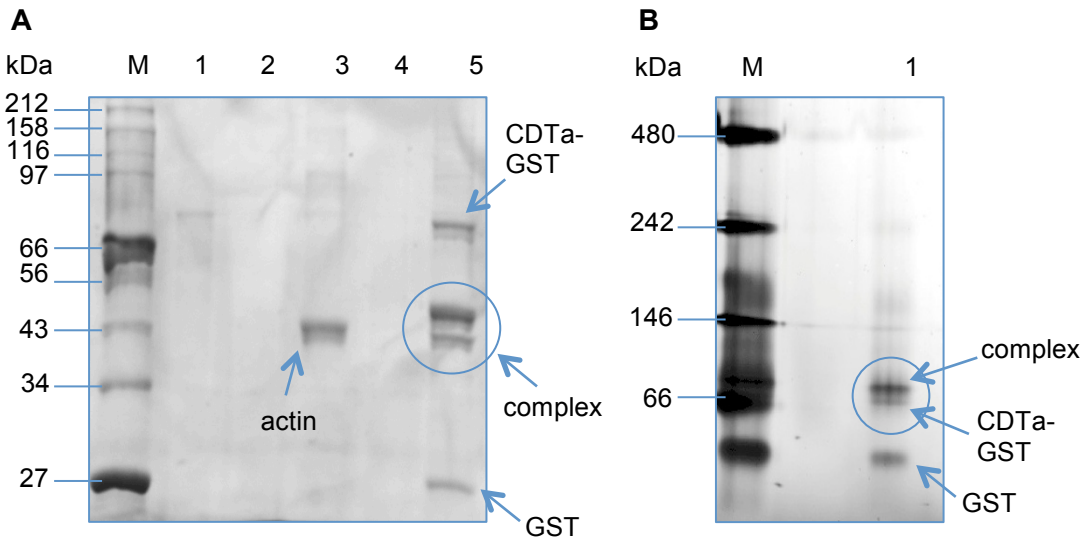


Figure 4.1.3j- Complex formation attempt 4b results. A: 10 % Tris-glycine SDS-PAGE displaying the steps to complex formation: Lanes 1 and 2 represent the CDTa-GST load and flow through. Lanes 3 and 4 represent the actin + IaγA + β-TAD load and flow through respectively. Finally lane 5 represents the final elution step. B: 4-16 % Tris-glycine Native gel displaying the final elution step in lane 1.

Firstly, the CDTa must be completely saturated with β-TAD and secondly, actin should be 100 % in its monomeric G-actin form, because the trimer could be either actin or CDTa based on the molecular weight of the trimer band, therefore lavγA was also increased to prevent polymerisation. In addition, the incubation period was extended to overnight rather than 1 hour, to ensure that all of the CDTa-GST bound to the column is saturated with actin for a long period to encourage complex formation.

Figure 4.1.3j-A displays the initial steps towards complex formation, which is significantly different from the previous attempt in a number of ways. Firstly, this time we have complete binding of CDTa-GST (lane 1) to the GST resin with no protein in the flow through (lane 2). Secondly, the amount of actin + lavγA + β-TAD (lane 3) washed off the resin (lane 4) is completely diminished suggesting all of the actin has bound CDTa thereby forming a complex. The final complex eluted from the column is displayed in lane 5, and the double band shares the same pattern as the previous attempt, except there is some CDTa-GST and GST contamination. Again to confirm the presence of a complex, the sample was analysed using a 4-16 % Native gel, with results displayed in Figure 4.1.3j-B. There is a clear band at approximately 80 kDa, which is a clear indication that a CDTa-actin complex has formed, however there is still the CDTa-GST band at approximately 70 kDa and some GST contaminant too at approximately 30 kDa. So it appears there has been some preliminary success in terms of CDTa-actin complex formation, however some refinement of the protocol is required to gain a higher yield for crystallisation trials, in addition to removal of some of the contaminants.

Formation of a CDTa-CDTb Complex

In attempts to form a CDTa-CDTb'' complex, a variety of ratios were tried to find the optimum for complex formation. In addition, the presence of a detergent (β-OG) was trialled in the hope that this would help to stabilise the CDTb'' oligomer. After combining all of the complex components and incubating for the set time length, the samples were all passed through a gel-filtration column. Each of the three attempts yielded the same chromatogram profile of two individual sharp peaks, see Figure 4.1.3k-A for an example.

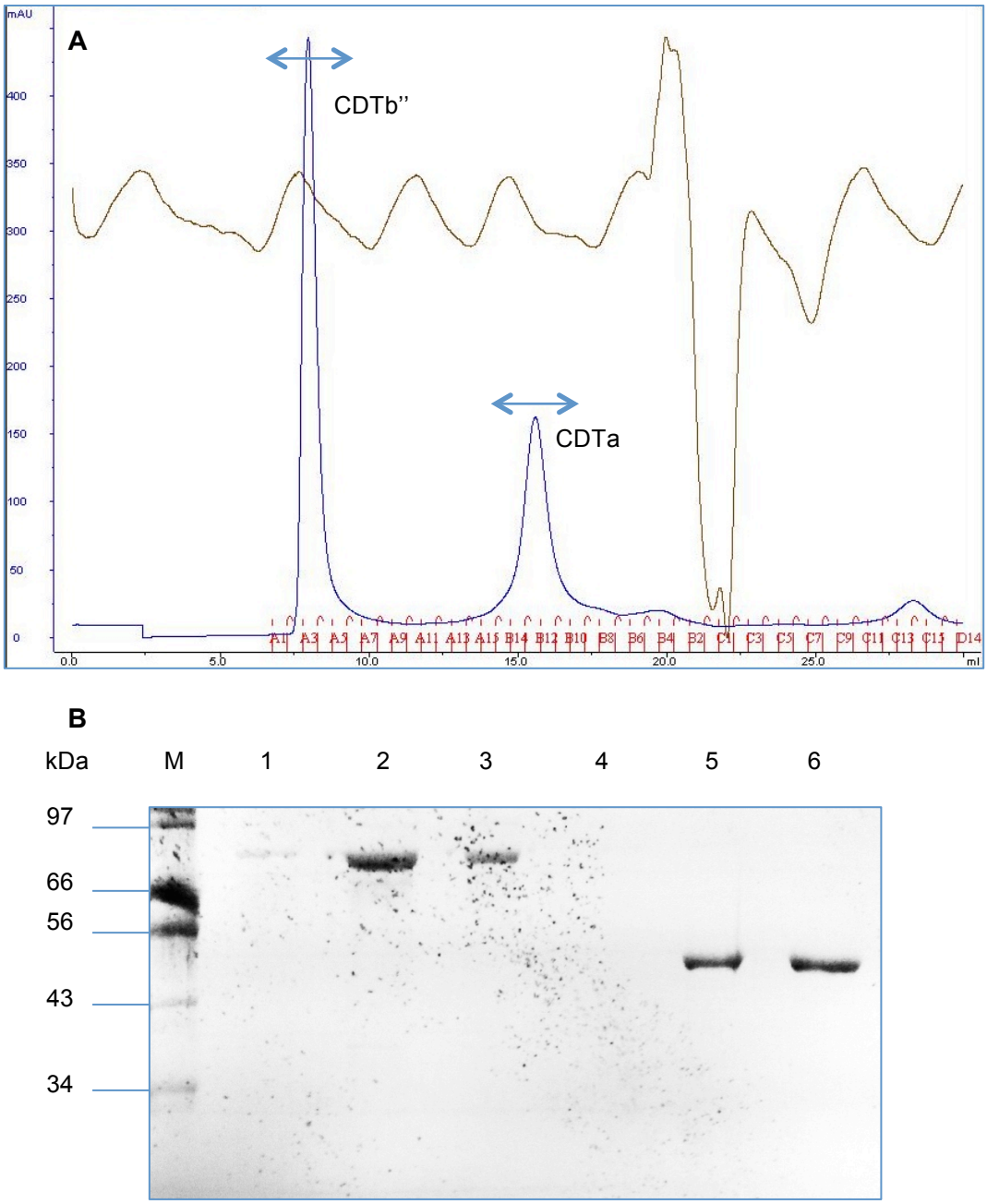


Figure 4.1.3k. CDTa-CDTb'' size-exclusion chromatography results. A: The chromatogram for the size-exclusion run clearly showing two distinct peaks in the blue Absorbance at 280nm trace, (brown is conductance). B: Fractions A2, A3, A5, B15, B13 and B11 were analysed by 10 % Tris-glycine SDS-PAGE.

The two peaks were analysed by 10 % Tris-glycine SDS-PAGE and unfortunately the first large peak contained only CDTb'' for all of the attempts and the second peak contained only CDTa for all of the attempts. This is a clear indication that complex formation was unsuccessful for all three attempts.

4.2 CDTa Mechanism

4.2.1 Introduction

CDTa Active Site

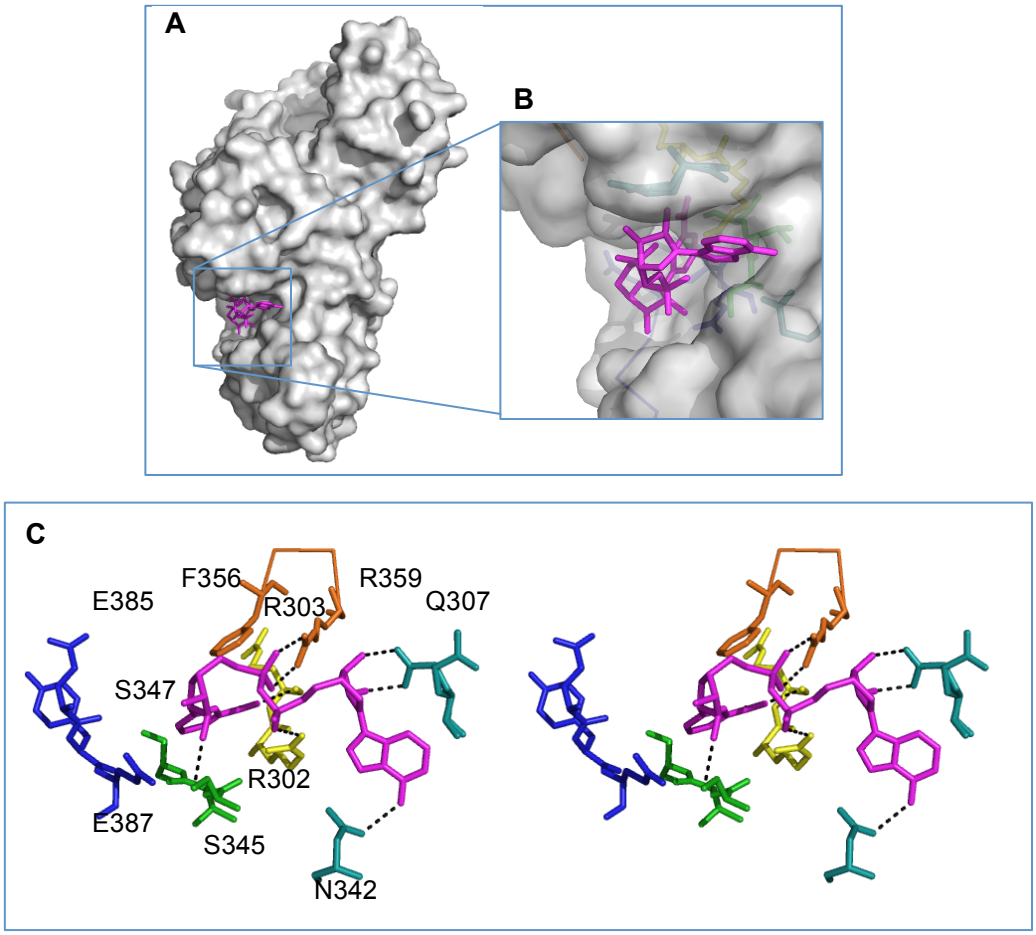


Figure 4.2.1a- The crystal structure and active site view of CDTa. A: Overall structure of CDTa displayed as surface (grey), with NAD bound displayed as sticks (magenta). B: Close up image of the active site cleft. C: Active site interactions displayed in stereo with the various ADPRT structural motifs coloured as follows: ExE-motif and ARTT-loop in blue, STS-motif in green, PN-loop in orange, Arg-motif in yellow and other residues involved in NAD binding in teal (Sundriyal *et al.*, 2009)**. Images rendered using PyMOL (Version 1.5.0.4 Schrödinger, LLC).**

The structure of CDTa was solved both with and without NAD bound at pH 9.0, in addition to pH 4.0 and 8.5 without NAD bound (Sundriyal *et al.*, 2009). The structure of CDTa with NAD bound can be viewed in Figure 1.2.2b, and the overall surface structure showing the active site cleft can be viewed in Figure 4.2.1a.

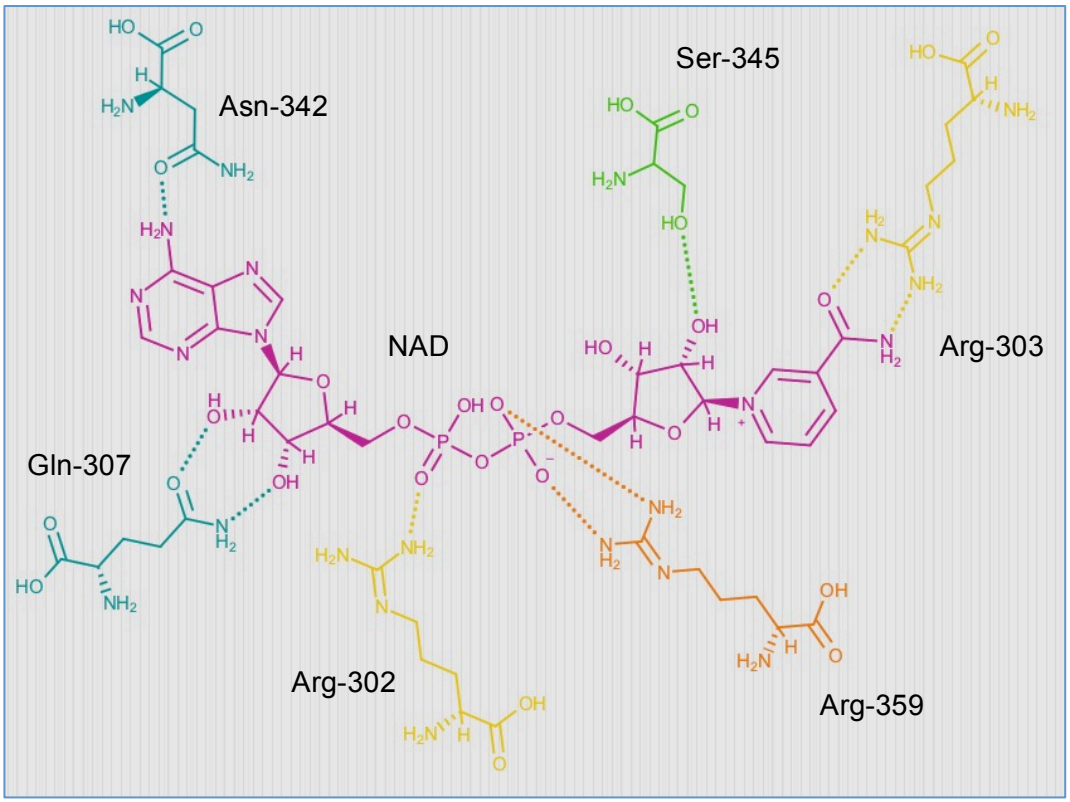


Figure 4.2.1b- Schematic representation of NAD binding in CDTa active site. Residues are colour coordinated with those in the previous active site figures and are not orientated as in active site but are placed for easy viewing of specific interactions.

The structure of CDTa shows that the NAD substrate binds tightly in the active site cleft with a number of key residues and ADPRT structural motifs contributing to its binding and stability. These interactions can be viewed in stereo mode, displayed in Figure 4.1.2a or schematically in Figure 4.2.1b.

The structural motifs that are characteristic of the ADPRT family are observed in the active site of CDTa including the ExE-motif and ARTT-loop (displayed in blue) which are key for ADPRT activity, the PN-loop and Arg-motif (displayed in orange and yellow respectively) and the STS-motif (displayed in green), which due to its interaction with NAD and the ExE-motif, may be involved in both ligand binding and catalysis (Sundriyal *et al.*, 2009). The ExE-motif has been shown to be directly involved in ligand binding and catalysis in other ADPRT family members such as the Ia toxin, however in this structure the ExE-motif is out of range for NAD binding, which is unexpected. This particular arrangement of Glu-385 and Glu-387 in the active site could be due to the movement of the flexible ARTT-loop (Gulke *et al.*, 2001; Mauss *et al.*, 1990).

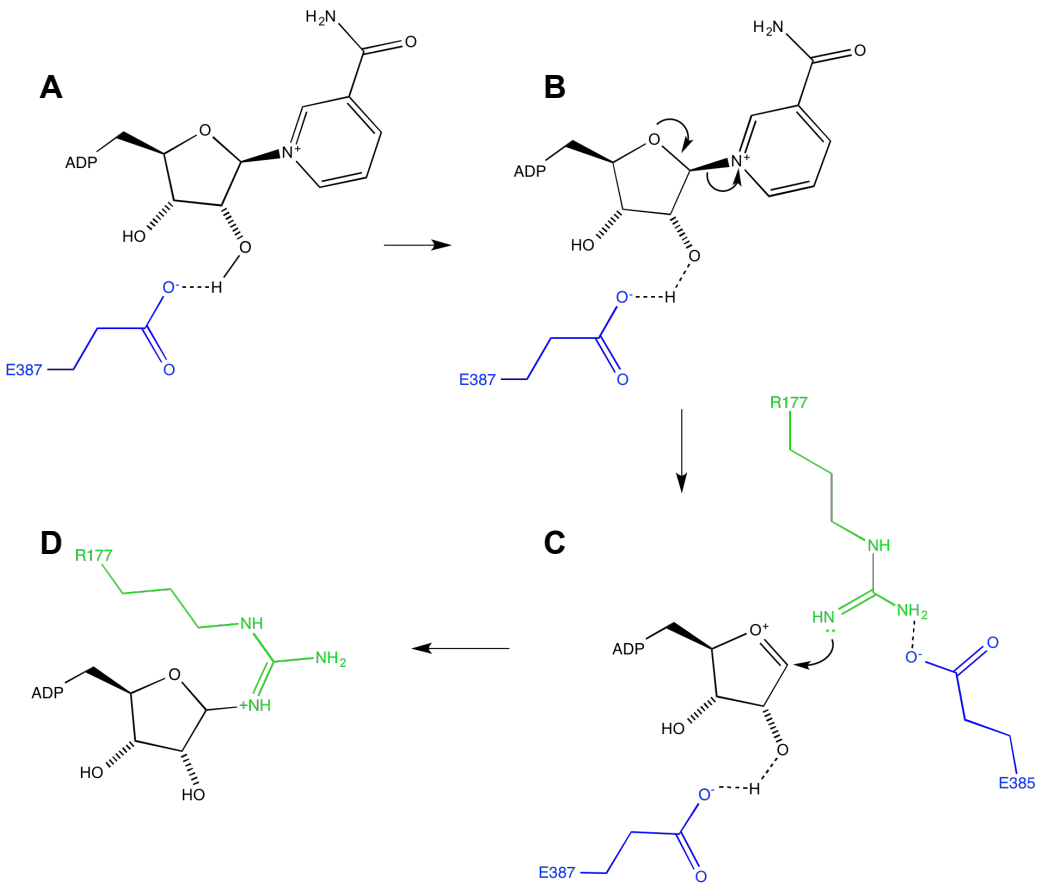


Figure 4.2.1c- Predicted mechanism for ADP-ribosylation of actin by CDTa. A: H-bond formation between E387 (blue) and 2 OH' of ribose (black). B: Loss of the nicotinamide group. C: E385 (blue) brings R177 (green) into active site and the ribose group oxocarbenium ion is susceptible to nucleophilic attack. D: R177 (green) is ADP-ribosylated by CDTa. Figure adapted from Davies *et al.* (Davies *et al.*, 2011)**. Image creating using ChemDraw (2007 PerkinElmer Informatics).**

Mechanism of ADP-Ribosylation

Using both the structural aspects of the CDTa active site along with mutagenesis studies on homologous ADPRT family members, the mechanism by which CDTa ADP-ribosylates monomeric actin can be predicted. The proposed mechanism is the SN1 mechanism via two intermediates; firstly an oxocarbenium ion intermediate and secondly a cationic intermediate. This is commonly thought to be the mechanism observed amongst the ADPRT family of toxins.

In the above SN1 mechanism, the catalytic glutamate Glu-387 in CDTa forms a H-bond with the 2'OH of ribose forming an oxocarbenium intermediate and rendering the ribose group vulnerable to nucleophilic attack. Following cleavage of NAD and formation of an oxocarbenium ion, it is thought that the ARTT-loop is rearranged to bring the Glu-385 residue to the reaction centre to stabilise the transfer of ADP-ribose to Arg-177 of actin, see Figure 4.2.1c.

Objectives

The aims of this part of this subchapter were to perform mutagenesis studies on some of the key active site residues of CDTa and determine their importance in the activity of CDTa, thereby testing the hypothesis of the SN1 mechanism. This should help resolve the issue of the importance of the ExE-motif in the catalytic activity of CDTa from the current contradicting evidence of positioning and involvement.

4.2.2 Materials and Methods

Site Directed Mutagenesis of Active Site Residues

Initial Attempts at Site Directed Mutagenesis

Mutagenesis of active site residues was performed at the DNA level using the pMAL-p2x-cdtA plasmid as a template and the respective oligonucleotides as displayed in Table 4.2.2a. For each mutant, two complementary synthetic oligonucleotides were required for the desired mutation. The mutated plasmids were transformed into *E.coli* Top 10 cells and the respective mutations were confirmed by DNA sequencing (Eurofins Genomics MWG Operon), resultant sequences were aligned with the native CDTa using T-Coffee, see Appendix 3 (Notredame *et al.*, 2000).

Mutation	Primers
S345A	5' - CCGAACTTTATTGCAACTAGT – 3' 5' - ACTAGTTGCAATAAAGTTCGG – 3'
S345F	5' - CCGAACTTTATTTTCACCAGC – 3' 5' - GCTGGTGAAAATAAAGTTCGG – 3'
S345R	5' - CCGAACTTTATTCGTACCAGC - 3' 5' - GCTGGTACGAATAAAGTTCGG – 3'
S345Y	5' - CCGAACTTTATTTACACCAGC – 3' 5' - GCTGGTGTAAATAAAGTTCGG – 3'
E385Q	5' - GTTATGCAGGTCAATATGAAGTG - 3' 5' - CACTTCATATTGACCTGCATAAC - 3'
E387Q	5' - AGGTGAATATCAAGTGCTTTTA – 3' 5' - TAAAAGCACTTGATATTCACCT – 3'
E387R	5' - GTGAATATCGTGTGCTTTTA – 3' 5' - TAAAAGCACACGATATTCAC - 3'

Table 4.2.2a- The synthetic oligonucleotides designed for specific mutations highlighted in yellow.

The successfully mutated plasmids were then transformed into *E.coli* BL21 (DE3) codon plus cells for expression trials. Expression was carried out as described in Chapter 4.1.2 under 'Pre-existing Protocols'. In brief, cells were grown in TB medium at 37 °C until they reached an OD_{600}= 0.6-0.8 whereby the temperature was then reduced to 20 °C and protein expression was induced with 1 mM IPTG and was harvested after 4 hours.

Optimised Mutation Production

As discussed in Chapter 4.1.3, the yield of pure CDTa obtained was increased by using the pGEX-6p1-cdtA plasmid to express the protein which enabled purification by GST-affinity chromatography. This alteration in the protocol improved the final yield of CDTa significantly. Therefore, to improve the yield of the mutants, the same technique was applied: individually the pMAL-p2x plasmids containing the mutated CDTa DNA were digested using BamH1 and Sal1 and were ligated into the pGEX-6p1 vector using T4 DNA ligase. DNA was sequenced to ensure successful ligation of mutants into the pGEX-6p1 vector.

Expression and Purification of CDTa Mutants

Expression of CDTa Mutants

Expression of CDTa mutants was performed using the optimised protocol from Chapter 4.1.2. *E. coli* BL21 (DE3) codon plus cells were transformed with the mutant pGEX-6p1-cdtA expression constructs. Overexpression of soluble mutants was obtained using 2 L baffled flasks containing 1 L of LB medium supplemented with 100 µg/ml ampicillin. The medium was inoculated with 10 ml overnight culture of the expression host, grown at 37 °C in LB supplemented with 100 µg/ml ampicillin. Once the culture reached OD_{600} = 0.8-1.0, the temperature was reduced to 20 °C. After the temperature reached 20 °C, 1 mM IPTG was added to induce expression of CDTa. The cultures were harvested after 4 hours, using 500 ml centrifuge tubes in a 10:500 rotor, spinning at 8,500 rpm for 10 min at 4 °C. The cell pellets were re-suspended in 2 ml/g of pellet of lysis buffer; 50 mM Tris-HCl pH 8.0 and 20 mM NaCl. The re-suspended pellets were then rapid frozen in liquid nitrogen and stored at -80 °C until further use.

Purification of CDTa Mutants

Purification of CDTa mutants was performed using the optimised protocol from Chapter 4.1.2. In brief: the mutants were purified using GST-tagged affinity chromatography in step 1. The GST-tag was then removed during dialysis (to remove reduced glutathione) using PreScission Protease (GE Healthcare) and

non-tagged protein was passed back through the GSTrap column (GE Healthcare) and was collected in the flow through. This was repeated to ensure complete removal of the cleaved GST-tag. The final pure mutants were then concentrated to 0.5 mg/ml using 10 kDa cut-off concentrators (Amicon Ultra 15, Millipore), and were rapid frozen in liquid nitrogen and stored at -80 °C, until further use.

Effect of Active Site Mutations on Catalytic Activity of CDTa

Vero Cell Assays

To assess the cytotoxicity of the mutated CDTa, Vero cell assays were performed using a previously established protocol with some modifications (Sundriyal *et al.*, 2010). The cells chosen were African green monkey kidney epithelial cells (Vero cells) as they have previously shown cell death in the presence of CDTa and CDTb' combined, as well as Toxin A and B. Vero cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) containing 10 % fetal calf serum (FCS) and 2 mM glutamine at 37 °C in the presence of 5 % CO_2. Cells were routinely trypsinized and used to coat 96 well plates in a volume of 200 µl (10^4 cells/well) of complete DMEM (supplemented with 10 % FCS and 2 mM glutamine) and were incubated for 24 hours to allow formation of a confluent monolayer. Prior to any assays, the medium was carefully removed and the cells were washed twice with Dulbecco's phosphate buffered saline (DPBS) and the cells were incubated for an hour in 100 µl DMEM (supplemented with 10 % FCS and 2 mM glutamine). A series of negative control assays were performed whereby Vero cells were incubated for 24 hours with 100 µl volumes of serum free DMEM, CDTa buffer and CDTb' buffer (separately), 250 ng CDTa and 250 ng trypsin activated CDTb' (separately) to ensure no cell death under 'non-cytotoxic' conditions. In addition Toxin A and Toxin B positive controls were performed to ensure that the Vero cells were responsive to the standard cytotoxicity assay (established by PHE). Cells were incubated for 24 hours in 100 µl of 50 ng/ml Toxin A and 0.5 ng/ml Toxin B, at 37 °C in the presence of 5 % CO_2. All of the assays were performed in triplicate. Alongside the control experiments, 250 ng of both CDTa and trypsin activated CDTb' were added to the cells in 100 µl to confirm the cytotoxic activity of CDTa and CDTb' combined.

Then 250 ng of each mutant was combined with 250 ng of trypsin activated CDTb' individually (in 100 µl) and were incubated alongside the CDTa+CDTb' control for 24 hours at 37 °C in the presence of 5 % CO_2. The cells were examined after 24 hour for evidence of cytotoxic effects. Assays were performed with the help of Dr Joanna McGlashan and Dr April Roberts at Public Health England, Porton Down.

Western Blot Assays

In order to confirm whether or not the mutations were affecting the binding of NAD or the transfer of ADP-ribose to actin, a western blot was performed using biotinylated-NAD (biotin-NAD) as the substrate. Samples were subjected to 10 % Tris-glycine SDS-PAGE and streptavidin horseradish peroxidase (strep-HRP) conjugate was used to detect biotin-NAD. Preliminary attempts were made at optimising the assay by using a range of CDTa concentrations, biotinylated-NAD concentrations and varying ratios of strep-HRP conjugate to develop the blot. Negative controls were performed including CDTa positive, biotinylated-NAD negative, and biotinylated-NAD positive, CDTa negative, to confirm that neither biotinylated-NAD nor CDTa individually alone would be detected in the blot.

The final protocol for the Western blot assay is as follows: CDTa and biotin-NAD were mixed in equal molar ratio, whilst actin was mixed in equal molar ratio with latA, which is a small molecule that maintains actin in its monomeric form, preventing polymerization. Actin was then added to the CDTa-biotin-NAD mix in a 1:1 molar ratio. The reaction was incubated at room temperature with gentle rocking for 1 hour. The same was repeated for each of the seven purified mutants. The samples were treated with SDS-PAGE sample loading dye and were immediately loaded onto a 10 % Tris-glycine SDS-PAGE, without heating the samples. The SDS-PAGE was run at 200 V for 1 hour, after which the samples were transferred to nitrocellulose membrane using standard techniques for 1 hour at 50 V. The membrane was blocked overnight in 1 % bovine serum albumin (BSA) in phosphate buffered saline pH 7.4 containing 0.05 % Tween-20 (PBST). The membrane was then washed for 5 min twice in PBST, after which the membrane was incubated with 1:1000 strep-HRP in PBST for 1 hour. The membrane was then washed for 5 min, 6 times before the developing solution

was added: 50 mM Tris-HCl pH 7.5, 0.03 % H_2O_2 and 0.1 % 3, 3'-Diaminobenzidine (DAB). Developing solution was removed immediately after bands appeared on the gel to prevent over exposure.

Structure of CDTa Mutants

To complement the assay work, the plan was to solve the structure of the mutants in order to visualize exactly what effect the mutants have on the structural arrangements in the active site. The CDTa native structure is available already; therefore this could be used as a model for molecular replacement to solve the phase problem. Therefore the mutants were concentrated to 4 mg/ml which is the concentration used for crystallisation of the native, and 24-well hanging drop plates were performed using the following crystallisation condition: 0.1 M malonate imidazole boric acid buffer (MIB) pH 9.0 and 20 % PEG 1500 (Sundriyal *et al.*, 2009). The crystallisation screens were set up in triplicate for each mutant and the plates were incubated at 16 °C, and monitored weekly for crystal growth. Until date there are no signs of crystal growth for any of the mutants, however crystallisation screens were performed towards the end of the project and are still regularly checked for signs of crystal growth.

As the original protocol used to solve the CDTa structure involved expression of protein from the pMAL-p2x-cdtA construct, the protein was purified using a different technique to that which was used for the mutants. The mutants were purified using GST-affinity chromatography and were expressed from a pGEX-6p1 construct. Therefore in an attempt to more accurately replicate the original CDTa crystallisation conditions, the individual mutants were buffer exchanged into the final CDTa buffer used in the original purification protocol: 50 mM Tris-HCl pH 8.0 and 20 mM NaCl. Buffer exchange was performed using small volume dialyzing tubes (D-tube dialyzer midi MWCO 6-8 kDa, Millipore) overnight at 4 °C to ensure complete transfer into the CDTa buffer. The above crystallisation trials were then repeated using the original crystallisation condition, concentration and temperature, with the hanging drop vapour diffusion method. Plates were monitored daily for the first 7 days, then weekly thereafter.

4.2.3 Results

Site Directed Mutagenesis of Active Site Residues

As discussed in the materials and methods section (Chapter 4.2.2), the original protocol used to obtain pure CDTa was significantly altered to improve the final yield. The initial and most key change in protocol was that the original CDTa DNA was cloned into an alternative vector (pGEX-6p1). As the mutants were created prior to this change in protocol, the mutants were also cloned into pGEX-6p1 so that they could be expressed and purified in an identical manner to the native CDTa.

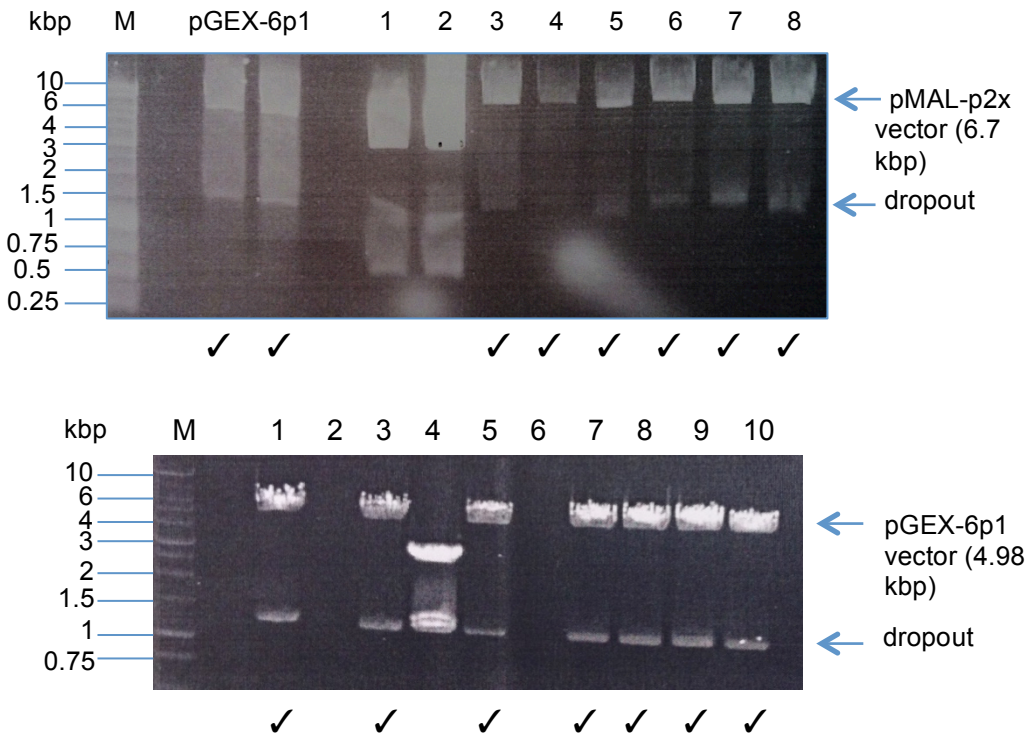


Figure 4.2.3a- 0.8 % agarose gels showing examples of the various stages of cloning for the various CDTa mutants. The upper gel shows the first step in cloning whereby the pMAL-p2x vector containing the mutated DNA was digested with BamH1 and Sal1, alongside digestion of the pGEX-6p1 vector with the same restriction enzymes. The lower gel shows the quick digest checks performed after ligation to ensure the presence of the CDTa mutant dropouts from the pGEX-6p1 vector. The ticks indicate successful digests. The pMAL-p2x vector is approximately 6.7 kbp long as indicated in the upper gel and pGEX-6p1 is 4.98 kbp as indicated by the shift downward in the vector band.

The mutant plasmids (pMAL-p2x-mutant) were digested alongside the new vector pGEX-6p1 using BamH1 and Sal1 restriction enzymes, see Figure 4.2.3a upper gel for an example. The CDTa mutant DNA dropouts were extracted from the gels and ligated into the digested pGEX-6p1 vector using T4 DNA ligase. A quick digest of the ligated mutants was performed, with an example of some of the results in Figure 4.2.3a, lower gel, to show the dropout in the new vector, indicating successful cloning.

Expression and Purification of CDTa Mutants

Expression of CDTa Mutants

The CDTa mutants were expressed using the optimised protocol from Chapter 4.1.2, using BL21 (DE3) codon plus *E.coli* cells.

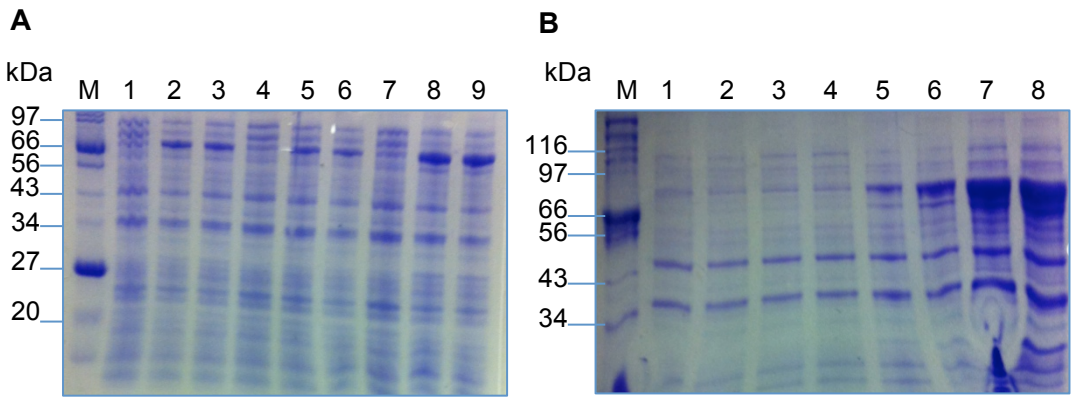


Figure 4.2.3b- Soluble expression of CDTa mutants. A: 12 % Tris-glycine SDS-PAGE displaying mutant expression: lanes 1, 4 and 7 are pre-induced samples, lanes 2 and 3 are S345R post-induction samples (duplicate), lanes 5 and 6 are S345Y post-induction samples (duplicate) and finally lanes 8 and 9 are E385Q post-induction samples (duplicate). B: 10 % Tris-glycine SDS-PAGE displaying mutant expression: lanes 1-4 are pre-induced samples and lanes 5, 6, 7 and 8 are post-induction samples of S345A, S345F, E387Q and E387R respectively.

There is a notable increase in the intensity of the bands at 77 kDa following induction of expression with 1 mM IPTG. This suggests the mutants are being expressed in a similar manner to CDTa. There are notable differences in the expression levels across the various mutants, but there are often batch-to-

batch variations in expression levels for the same protein as a result of variation in OD_{600} or temperature fluctuations during cell growth.

Purification of CDTa Mutants

The CDTa mutants were purified using the optimised protocol designed for the native CDTa. After an initial GST-affinity chromatography step, the GST-tag was cleaved off using PreScission Protease. The cleaved protein was then passed back through the GSTrap column multiple times collecting the mutant protein in the flow through. This allowed removal of the GST-tag by allowing it to rebind to the column, see Figure 4.2.3c, for step-by-step analysis by SDS-PAGE.

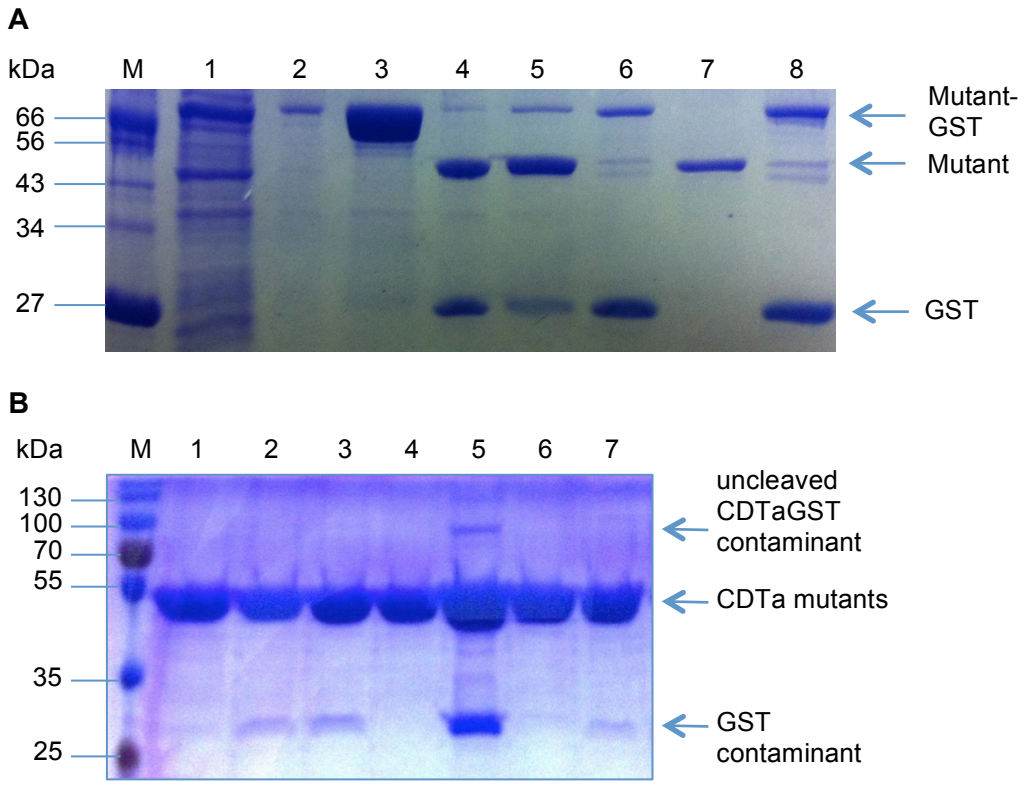


Figure 4.2.3c- Purification of CDTa mutants with each step analysed by 10 % Tris-glycine SDS-PAGE. A: Lanes 1 and 2 represent the protein loaded onto the column and the flow through respectively. Lanes 3 and 4 represent the eluted protein and the cleaved protein respectively. Lanes 5 and 6 represent the 'pure' protein flow through and the eluted GST tag as indicated. Lanes 7 and 8 represent the second purification step where 7 is the pure CDTa flow through and 8 is the remaining GST and uncleaved protein eluted from the column. B: Lanes 1-7 represent the final purified mutants S345A, S345F, S345R, S345Y, E385Q, E387Q and E387R respectively.

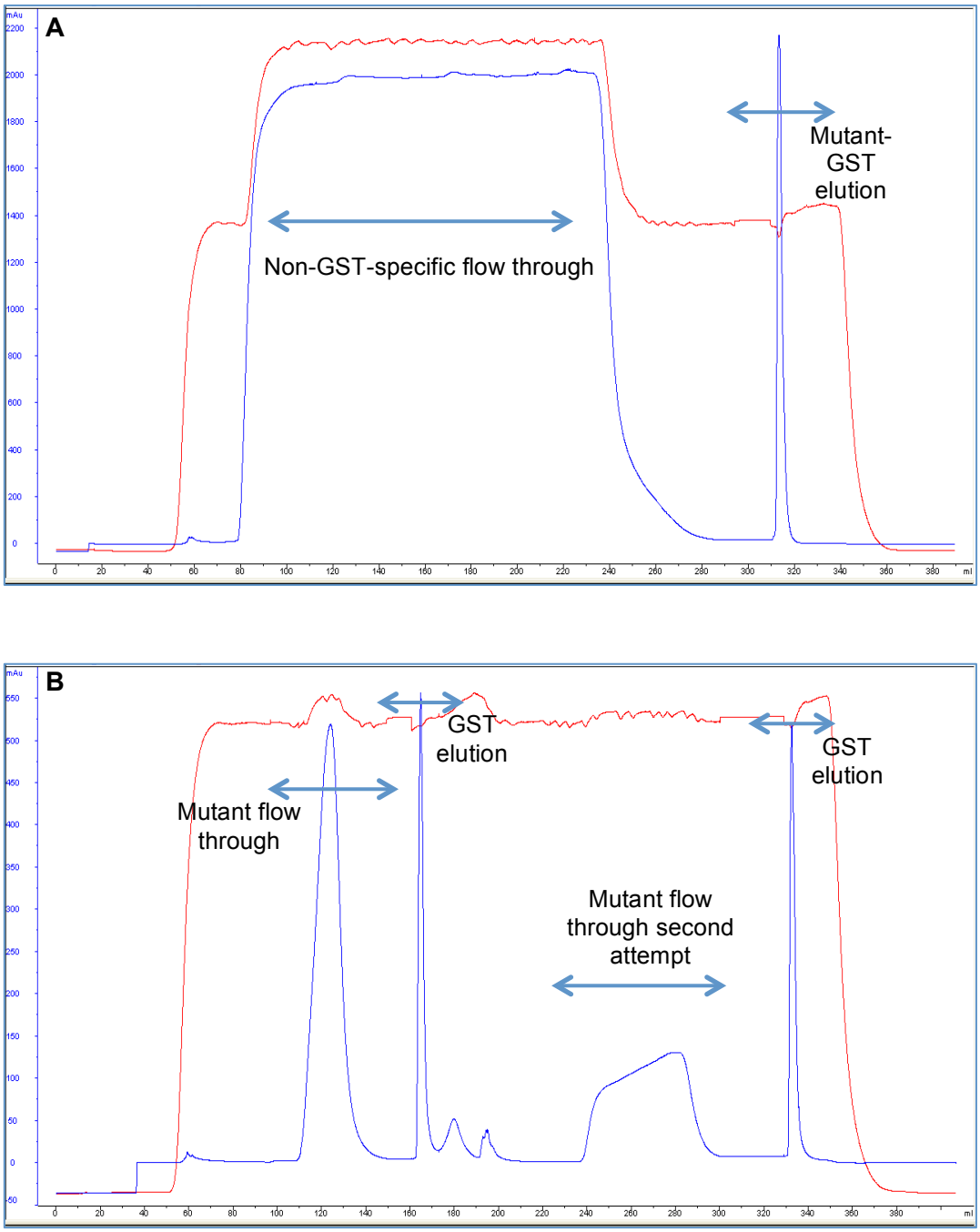


Figure 4.2.3d- Example of Purification of CDTa Mutants by GST-affinity Chromatography. The blue lines represent the absorbance at 280 nm and the red line represents conductivity. A: First step, cell lysate is applied to GSTrap column allowing CDTa-GST to bind and non-specific proteins flow through. Bound CDTa-GST is then eluted using reduced glutathione. B: After cleavage of GST-tag, the second step involved, passing cleaved protein back through GSTrap column collecting the 'pure' CDTa mutant in flow through, whilst the GST-tag binds column. This is repeated twice to ensure complete removal of GST-tag.

Effect of Active Site Mutations on Catalytic Activity of CDTa

Vero Cell Assays

To assess the cytotoxicity of the CDTa mutants, Vero cell assays were performed. Vero cells were grown in DMEM containing 10 % FCS and 2 mM glutamine at 37 °C in the presence of 5 % CO_2. Cells were routinely trypsinized and used to coat 96 well plates. A series of negative control assays were performed whereby Vero cells were incubated for 24 hours with serum free DMEM, CDTa buffer and CDTb' buffer, 250 ng CDTa and 250 ng trypsin activated CDTb' (separately) to ensure no cell death under 'non-cytotoxic' conditions. These results are displayed in Figure 4.2.3e A-E respectively. As shown in these images, there is no cell death observed, confirming the Vero cells are not sensitive to either of the protein buffers or CDTa and trypsin activated CDTb' individually.

In addition Toxin A and Toxin B positive controls were performed to ensure that the Vero cells were responsive to the standard cytotoxicity assay (established by PHE). Figure 4.2.3e F and G display the Toxin A and Toxin B positive controls respectively. It is clear that there is complete cell death in the presence of both Toxin A and B individually (cell death indicated by rounding of cells). Cells were then incubated for 24 hours in the presence 250 ng of both CDTa and trypsin activated CDTb' to test the effect of the combination of both binary components. The results are displayed in Figure 4.2.3e H, where it is clear there is some cell death, but not to the same extent as Toxin A and B, but this is to be expected as these are known to have a clear cytotoxic impact on Vero cells. These results are consistent with those previously reported (Sundriyal *et al.*, 2010).

The Vero cells were then incubated in the presence of 250 ng of each mutant combined with 250 ng of trypsin activated CDTb' individually. The results for the S345A mutation are displayed in Figure 4.2.3e I, which shows a significant decrease in cytotoxicity but there is still noticeable cell death indicating the S345A mutation does not diminish the cytotoxic effect of CDTa. Images J-O in Figure 4.2.3e represent the remaining six mutants: S345F, S345R, S345Y, E385Q, E387Q and E387R, respectively. These remaining images show no signs of Vero cell death, which indicates that these active site mutations completely diminish the activity of CDTa preventing ADP-ribosylation of G-actin.

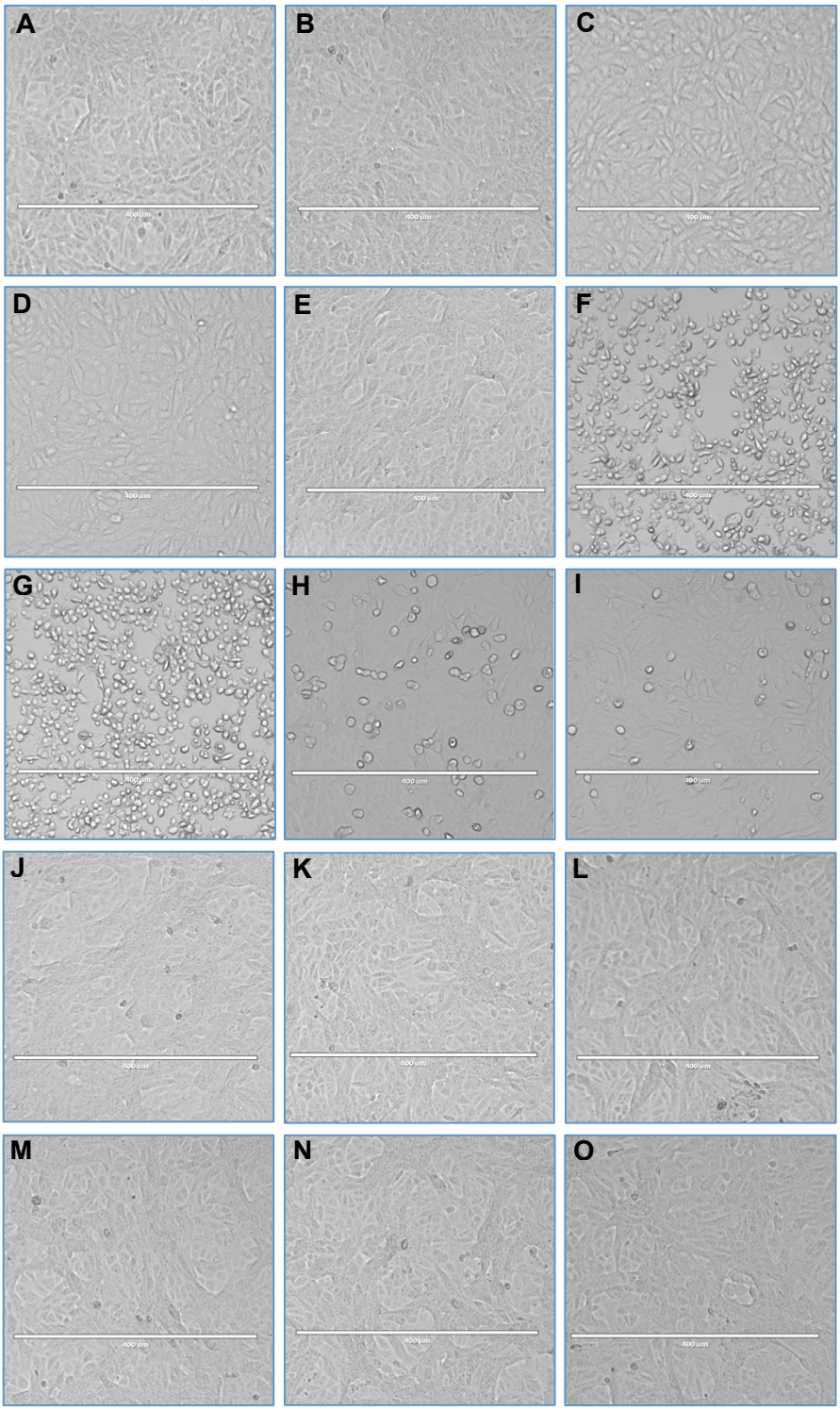


Figure 4.2.3e- Vero cell assay results. **A:** Media control. **B:** CDTa buffer control. **C:** CDTb control. **D:** 250 ng CDTa. **E:** 250 ng CDTb'. **F:** 50 ng/ml Toxin A. **G:** 0.5 ng/ml Toxin B. **H:** 250 ng CDTa + 250 ng CDTb'. **I:** 250 ng S345A + 250 ng CDTb'. **J:** 250 ng S345F + 250 ng CDTb'. **K:** 250 ng S345R + 250 ng CDTb'. **L:** 250 ng S345F + 250 ng CDTb'. **M:** 250 ng E385Q + 250 ng CDTb'. **N:** 250 ng E387Q + 250 ng CDTb'. **O:** 250 ng E387R + 250 ng CDTb'. The scale bar in each image represents 400 µm.

Western Blot Assays

Cell based assays were performed to assess the effect of the mutants using Vero cells. In addition, Western blots were performed to show more specifically how the mutations affect the activity of CDTa in terms of either binding of NAD to CDTa or the transfer of ADP-ribose to actin.

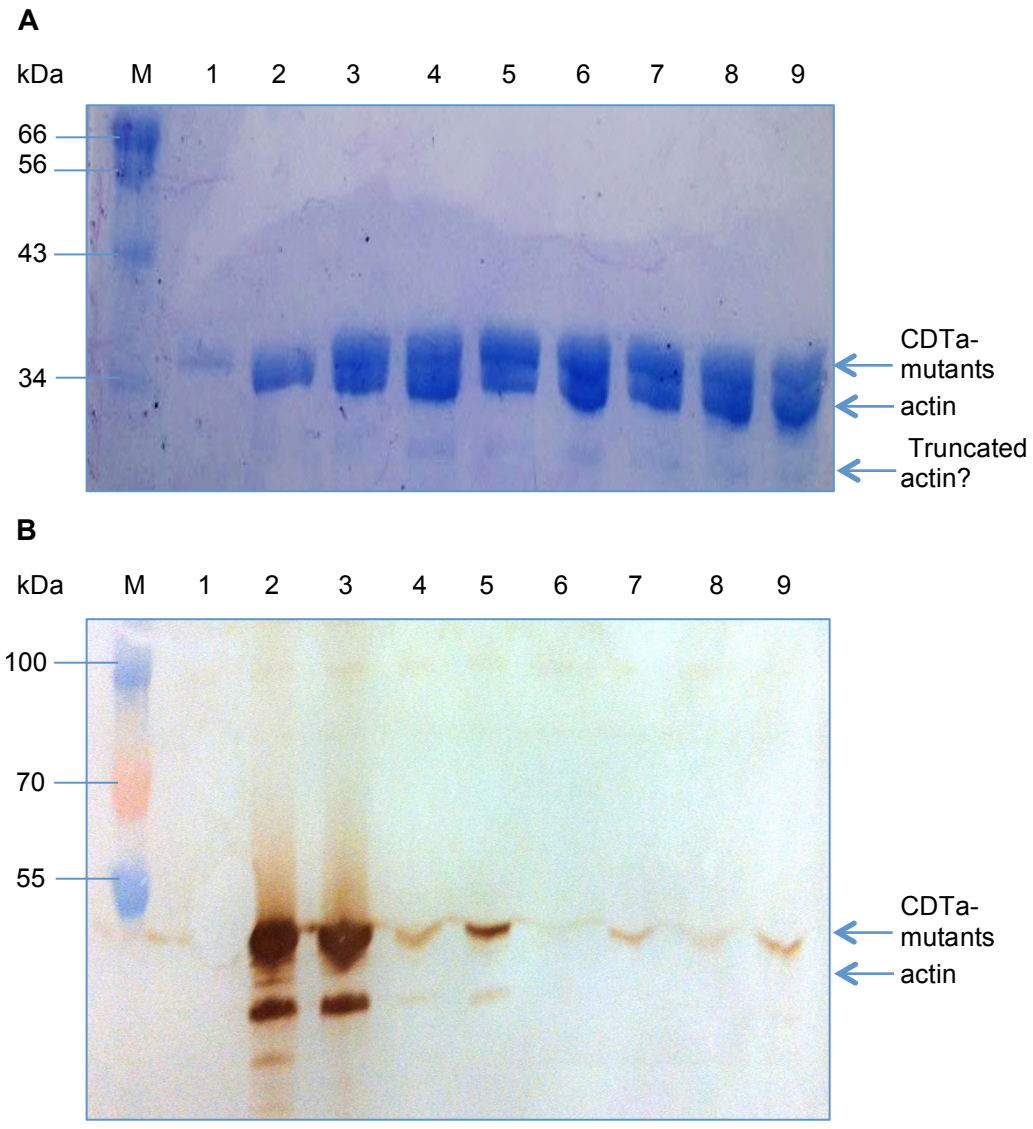


Figure 4.2.3f- Western blot assay results for CDTa and CDTa mutants. A: 10 % Tris-glycine SDS-PAGE gel picture displaying the protein bands transferred to the cellulose membrane for blotting. Lane 1 represents CDTa + biotin-NAD and lane 2 represents CDTa + actin + biotin-NAD. The remaining lanes 3-9 represent the mutations + actin +biotin-NAD in the order of S345A, S345F, S345R, S345Y, E385Q, E387Q and E387R respectively. B: The western blot developed with DAB displaying strep-HRP bound to biotin-NAD where present on the membrane. The lanes correspond with gel picture A.

Displayed in Figure 4.2.3f are the 10 % Tris-glycine SDS-PAGE results (A) alongside the Western blot results (B), which correspond to each other directly in terms of samples loaded. The first two lanes represent the control samples, CDTa + biotin-NAD, and CDTa + actin + biotin-NAD, which will confirm that the native binds NAD and also transfers the ADP-ribose group to actin. To confirm, the NAD is biotinylated at the ADP-ribose end and so can be detected after ADP-ribosylation. There is a faint band in the first lane which shows that CDTa has bound biotin-NAD and also in the second lane there is a much stronger band and a secondary band below which clearly indicates that biotin-NAD is present in both CDTa and actin, confirming the activity of the native protein. There is one slightly unusual point to note on both the SDS-PAGE gel picture and the Western blot picture and that is the presence of a faint lower molecular weight band. The close proximity stronger double band is characteristic of CDTa and actin when loaded on a gel together as they have similar molecular weights. It is possible that the lower molecular weight band is a truncated version of actin as it is not present in the first lane, which is the only lane that does not contain actin. The other evidence that points towards this theory is the very strong band of CDTa in lane 2 compared with lane 1. Both lanes contain an equal quantity of CDTa, yet in lane 1 there is a very faint Western blot band, but lane 2 is very strong. As there is no substrate for CDTa to transfer ADP-ribose to in lane 1, the biotin-NAD could possibly have been released, where as in lane 2 it is transferred to actin, which is why the band remains so strong. There is no biotin-NAD wasted in lane 2, hence the stronger band. As the lower molecular weight 'unknown' band appears highlighted in the Western blot in lane 2, this suggests that it is a substrate for biotin-NAD. The lower band is only present in the lanes containing actin, and this molecular weight band is not present in the SDS-PAGE analysis of the purified mutants, therefore, it may be a truncated or degraded actin band.

The following lanes 3-9 represent the mutants and the effect they have on NAD binding and/or ADP-ribose transfer to actin. It is clear from the Western blot that lane 3 shows the most consistent effect as the native protein, and this mutation is S345A. These results correlate to the Vero cell assays in which S345A was the only mutation that causes any cytotoxic effect on the cells. Lanes 4 and 5 that represent S345F and S345R respectively show strong CDTa binding and potentially a very small fraction of biotin-NAD present in the low molecular

weight band proposed to be actin. Lane 6 represents S345Y, which has very little biotin-NAD present either in CDTa or actin, which could mean this mutation prevents any biotin-NAD from binding CDTa at all. Based on the size of the residue that has replaced the existing one, it is more likely that the mutation caused a significant reduction in binding and cytotoxic efficiency. Lanes 7-9 represent E385Q, E387Q and E387R respectively and each displays a similar effect. The clear picture here is the complete lack of ADP-ribosylation of actin, but the presence of low volumes of biotin-NAD in the CDTa-mutant themselves, which indicates these mutations don't inhibit NAD binding but do inhibit ADP-ribosylation of actin.

Crystallisation of CDTa Mutants

Crystallisation trials were performed for all seven CDTa mutants, following the exact conditions used to crystallise the native CDTa including concentration, temperature and crystallisation solutions. This protocol had been used previously to successfully grow crystals of CDTa which were used to solve the structure at high resolution (Sundriyal *et al.*, 2009). Initially the trials were performed with the mutants in their final purification buffer, however after no success in crystal growth, the mutants were buffer exchanged into the buffer that was used to purify the native protein prior to crystallisation. Unfortunately, to date there have been no signs of crystal growth for any of the CDTa mutants.

4.3 Discussion

CDTa-Actin

One of the key aims of this chapter was to determine the 3D-structure of CDTa in complex with its target, actin. Theoretically, if the structure of this complex is understood, then we would be able to visualize the interactions between the two proteins that are necessary for ADP-ribosylation to take place. Understanding these interactions is key for prevention of actin ribosylation. Once the key interactions are resolved, it may be possible to design highly specific small molecules that can prevent this interaction, thereby inhibiting actin ribosylation, ultimately preventing cell rounding and cell death.

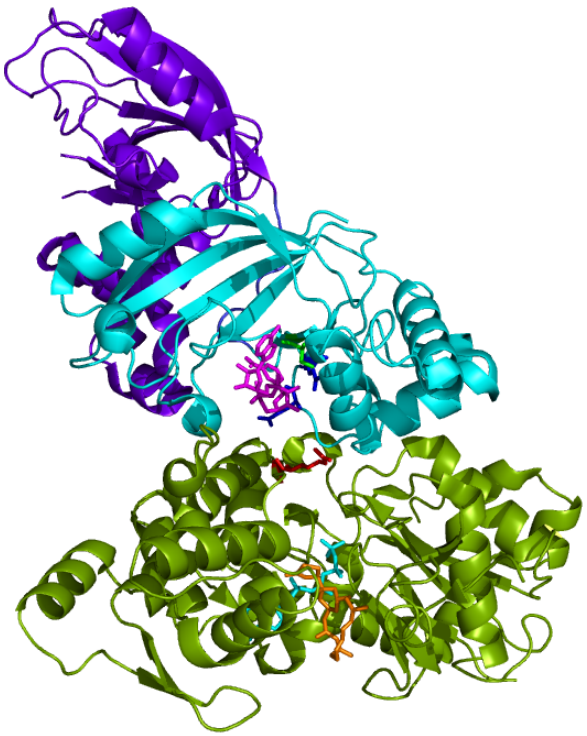


Figure 4.3a- The predicted structure of the CDTa-actin complex. The Ia-Actin structure was used for homology modelling, whereby the structure of CDTa was superposed onto that of Ia (Tsuge *et al.*, 2008)**. CDTa is displayed in cartoon with the N-terminal domain in purple and the C-terminal domain in cyan, corresponding to previous figures. The key active site motifs involved in binding NAD are coloured as previously: ExE-motif in blue and STS-motif in green. NAD is displayed as sticks in magenta. Actin is displayed in green as cartoon with the Arg-177 residue displayed in red sticks, ATP bound in cyan sticks and latrunculin A bound in orange sticks.**

Multiple attempts were made at CDTa-actin complex formation with some small-scale success for attempts 4a and 4b. A small amount of complex was successfully formed, however, optimisation is required. The main concept of the protocol was based on that used for the Ia toxin which is homologous to CDTa and is structurally almost identical (Tsuge *et al.*, 2008), see Figure 4.1.1c-A showing Ia toxin structure superimposed onto the CDTa structure. This protocol established for the Ia-actin complex was performed on a much larger scale and was a simple case of adding all of the components together and setting up crystallisation plates without any confirmatory step. However, this protocol was not applicable to CDTa, which could be for a number of reasons: one explanation could be that the rate of CDTa crystal growth is more rapid than that of Ia and CDTa may start forming microcrystals prior to complex formation. Alternatively, CDTa may be folded slightly differently when purified using the optimised protocol, as the protocol has been significantly adapted from the original. In order to visualise what the structure of the CDTa-actin complex may look like the Ia-actin complex was used for homology modelling, see Figure 4.3a.

CDTa-Actin Future Work

In order to obtain the CDTa-actin structure, the next step should be to scale up the 'on column' complex formation protocol from 'attempt 4b' as this was the most successful attempt to date. This could potentially provide enough complex to set up crystallisation screens. In addition to the standard screens, the Ia-actin crystallisation condition should be attempted using both the CDTa crystallisation condition and the Ia-actin complex crystallisation condition.

CDTa-CDTb

Another aim of this chapter was to determine the 3D-structure of CDTa in complex with the oligomeric transport component. For similar reasons to that of the CDTa-actin complex: the main aim behind structural determination of these complexes is to understand their interactions. If small molecules can block the interactions between the complexes, then the cytotoxic effects of CDTa can be reduced indirectly. Toxicity can be eliminated either by inhibiting the access to the cells via the interaction with CDTb or by inhibiting the collapse of the cytoskeleton by blocking the interactions with actin.

Attempts were made at CDTa-CDTb'' complex formation, whereby the two components were combined, incubated and subjected to size-exclusion chromatography. However, there was no success, which could be due to a number of reasons. There are two key theories behind the lack of success in formation of a complex:

- First and foremost, the way in which CDTa and CDTb'' interact with each other, and at what stage they interact is yet to be understood, and in addition, the way in which CDTb'' interacts with the target cell surface is not fully understood. It is thought that CDTb' must be activated by trypsin cleavage after which it either forms oligomers and interacts with the cell-surface receptors, or the monomers interact with the cell-surface receptors and slowly form oligomers (Barth, 2004). Once the CDTb'' has formed interactions with the target cell surface, it is predicted that CDTa will then interact with CDTb monomers or oligomers, but not until they are on the cell surface. This could be one reason behind the lack of CDTa-CDTb'' complex formation. When CDTb interacts with target cell surfaces, regardless of whether or not it is in the oligomeric or monomeric form, it may undergo conformational changes allowing CDTa to bind. This mechanism of CDTa entry into cells is not fully understood, and the interaction of CDTa and CDTb may be key to understanding how entry is initiated.

- Secondly, there is some ambiguity as to the status of CDTb'' after purification, as discussed in detail in Chapter 3.3. Briefly, the CDTb'' pure protein is eluted from the gel-filtration column at a higher molecular weight predicted size than expected (> 500 kDa), which at first could be explained as being a slightly larger oligomer than the predicted heptamer, or potentially a dimer of heptamers. However, after extensive analysis of CDTb'' using a variety of structural and biochemical techniques, it is possible that CDTb'' is not in the correct state for potential complex formation, see Chapter 3.3. Even if the first point was not applicable, and that CDTa could interact with CDTb'' prior to interaction with the cell-surface receptors, it may not be possible based on the current state of the purified CDTb''.

CDTa-CDTb Future Work

In order to form a CDTa-CDTb'' complex, a number of challenges need to be overcome. The first hindrance that should be addressed is the CDTb'' oligomeric state. If there is to be a chance at CDTa-CDTb'' formation it is crucial that CDTb'' is in the correct state to interact with CDTa, see Chapter 3.3 for efforts to overcome this problem. The second option to consider is introducing an additional component to the complex that acts as the cell surface. Recently the lipolysis-stimulated lipoprotein receptor (LSR) was identified as the host receptor for the binary toxin CDTb component (Papatheodorou *et al.*, 2011). If the theory of CDTb'' altering conformation on cell-surface binding is correct, then incorporating this receptor with CDTb'' could help with CDTa-CDTb''-receptor complex formation.

CDTa Mechanism of ADP-ribosylation of G-Actin

The selected residues Glu-385 and Glu-387 of the ExE-motif were mutated, as this motif is known to play a strong role in ligand binding and catalysis. They directly correspond to residues Glu-378 and Glu-380 from the Ia toxin which in previous mutagenesis studies were shown to play key roles in the cytotoxicity of Ia (Perelle *et al.*, 1996). In addition the Ser-345 residue was also selected for mutagenesis studies due to its key role in binding Glu-387 and NAD. The planned mutations included, removing the residues by mutating to alanine, altering the charge of the residue and changing the size of the residue to see what effect each of these changes has on the cytotoxicity. Due to time restraints, not all of the mutations were achieved at both the DNA level and the expression/purification level. Those mutants that were taken from DNA through to purification and were therefore used in the assays include S345A, S345F, S345R, S345Y, E385Q, E387Q and E387R.

To test the effect of these mutations on the activity of CDTa, both *in vitro* Vero cell assays and Western blot assays were performed. The Vero cell assay is a visual qualitative assay that will provide a yes or no answer as to whether or not the mutations affect the cytotoxicity of CDTa. The Western blot assays were

designed to differentiate between CDTa binding of NAD and ADP-ribosylation of actin, and therefore will test the effect of the mutations on either binding or catalysis or both.

The Vero cell assay results displayed in Figure 4.2.3e show a series of images taken of the various control and experimental assays performed on the Vero cells on a 96-well plate. In the first five images Figure 4.2.3 A-E are the negative controls performed using standard DMEM, CDTa and CDTb' buffers, and then CDTa positive, trypsin activated CDTb' negative and trypsin activated CDTb' positive ,CDTa negative controls. There is no cell death observed amongst the negative controls as expected, which is confirmation that CDTa and CDTb' are non-toxic to Vero cells individually. The following two images, Figure 4.2.3e F and H are positive controls performed using Toxin A and B that are known to be toxic to Vero cells in low concentrations. In these two images there is conclusive evidence of complete cell death, confirming the cells are responsive to the standard toxicity assay (established by PHE). The most essential assay performed was the combination of CDTa and trypsin activated CDTb', see Figure 4.2.3e-H. This control is necessary to confirm that when CDTa and trypsin activated CDTb' are combined, there is noticeable cell death, confirming that both components of the binary toxin are required to cause Vero cell death, which is consistent with the literature (Sundriyal *et al.*, 2010). Trypsin activated CDTb' was used rather than CDTb'' as this shows greater cytotoxicity when combined with CDTa, as established previously by Sundriyal *et al*. The amount of cell death observed for the CDTa + CDTb' control is much less than that observed for the LCTs which is to be expected as they are considered to be more potent toxins. As the quintessential CDTa + CDTb' assay was successful, even though the observed cell death is low, it was deemed suitable enough to continue with the CDTa mutants.

Each of the mutants were individually combined with trypsin activated CDTb' at the same concentration and were added to the Vero cells concurrently. The results for the mutants are displayed in Figure 4.2.3e images I-O and represent S345A, S345F, S345R, S345Y, E385Q, E387Q and E387R. When directly comparing these images to the CDTa + CDTb' control, it is evident that mutants J-O display no indication of cell death at all, suggesting these mutations have had a direct inhibitory effect on the cytotoxicity of CDTa. Image I which represents S345A is the only mutant that still shows blatant cell death, albeit

less than that of the native, there is still evidence that this mutation hasn't inhibited the activity of CDTa but may have slightly reduced its efficacy.

Displayed in Figure 4.2.3f are the Western blot assay results, which were performed in order to elucidate the effect the mutations have directly on binding and/or catalysis. The first two lanes represent the control samples, CDTa + biotin-NAD and CDTa + actin + biotin-NAD, which confirms that the native binds NAD and also that ADP-ribosylation of actin occurs. The NAD is biotinylated at the ADP-ribose end and so can be visible both before and after ADP-ribosylation.

The Western blot assay confirms that CDTa ADP-ribosylates actin in the native form, as biotin is observed in both CDTa and actin bands in lane 2. This assay also shares similar results to the Vero cell assays in that the S345A mutant also shows evidence of ADP-ribosylation of actin, as displayed in lane 3. In addition, the remaining mutants displayed no evidence of ADP-ribosylation of actin; see Figure 4.2.3f lanes 4-9 which again correlate to the results observed for the Vero cell assays.

The potential slight design flaw in this assay is that biotin-NAD is hydrolysable hence why we can observe ADP-ribosylation of actin. Therefore, the amount of biotin-NAD present in the native CDTa and mutants is not the end point amount as the biotin-NAD will be hydrolysed and released from the active site. This will complicate the assay to a certain extent in that it may be more difficult to determine whether or not the mutants are affecting NAD binding. However, due to the presence of the actin, the assay will be conclusive as to whether or not the actin is ADP-ribosylated. If the actin is not ribosylated this could mean the mutants are not binding the NAD in the first instant or if they are that they are not able to catalyse the hydrolysis and transfer of ADP-ribose to actin. For the native, it is possible to see trace amounts of biotin-NAD present in the control without actin, which shows promise for determination of whether or not the mutants affect NAD binding.

The lanes 4-9 represent the mutants that show no indication of ADP-ribosylation and so based on the previous statement of the observation for the native, the remaining mutants will be directly compared for NAD binding. Lanes 4 and 5 that represent S345F and S345R respectively show strong similar NAD

binding to the native and potentially a very small fraction of biotin-NAD present in the low molecular weight band proposed to be actin. Lane 6 represents S345Y, which displays a very faint band, which could mean this mutation prevents NAD binding. Based on the size of the residue that has replaced the existing one, it is likely that the mutation caused a significant reduction in binding and cytotoxic efficiency. Lanes 7-9 represent E385Q, E387Q and E387R respectively and each displays a similar effect. The clear picture is the complete lack of ADP-ribosylation of actin, but the presence of low volumes of biotin-NAD in the CDTa-mutant themselves, which indicates these mutations don't inhibit NAD binding but may inhibit ADP-ribosylation of actin.

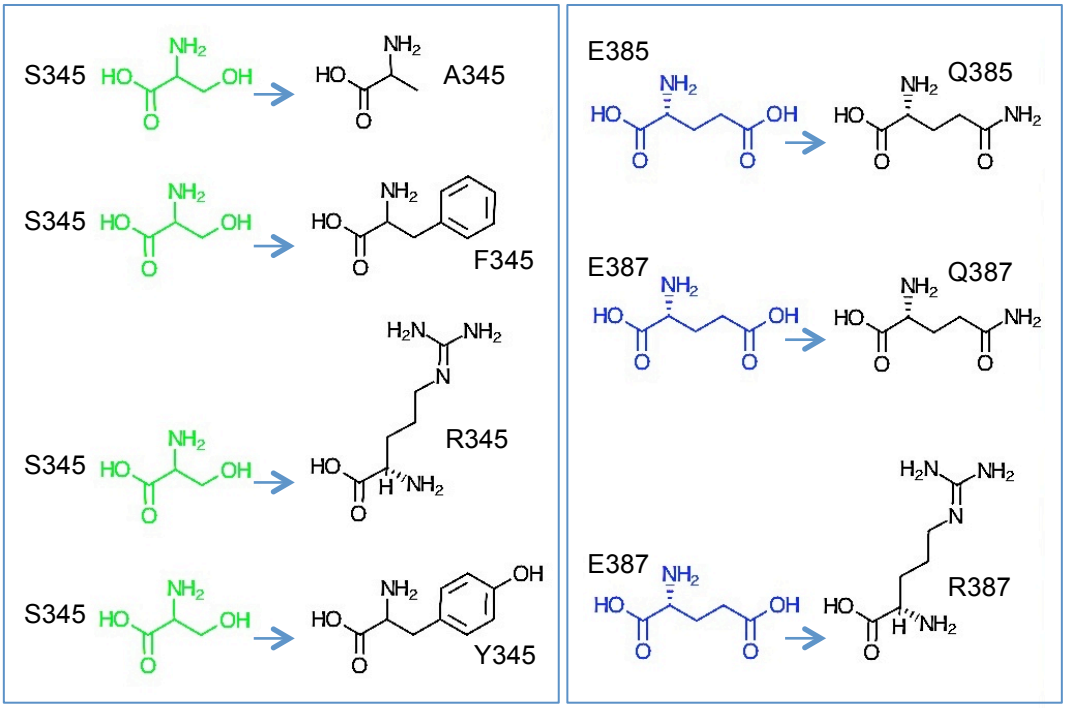


Figure 4.3b- Successfully performed mutations, expressed, purified and used in both Vero cell and Western blot assays. The Ser-345 residue was mutated to alanine, phenylalanine, arginine and tyrosine. Glutamic acid residues Glu-385 and Glu-387 were both mutated to glutamine, and Glu-387 was mutated to arginine.

Both the Western blot assay and the Vero cell assay have confirmed that S345A is the only mutation that maintains cytotoxicity. Figure 4.3b displays the mutations performed, and it is clear from the diagram that mutating a serine residue to alanine, only removes the –OH group, where as the mutations to either phenylalanine, arginine and tyrosine are much more significant in terms of

size and functional group properties. This would suggest that the functional –OH group of serine does not play an important role in the active site, but the size of serine does play a key role. As mutating this serine residue to alanine only marginally reduced the cytotoxicity but the other three mutations caused complete loss of cytotoxicity. This increase in size of the Ser-345 residue causes an adverse effect on cytotoxicity which could be due to a number of reasons, firstly it could prevent the residue from interacting with and stabilising Glu-387, or it could block the transfer of ADP-ribose to actin. The enlarging effect of the residue doesn't seem to affect NAD binding as demonstrated by the Western blot assay, with the exception of S345Y which could be purely due to space issues in the active site, reducing the NAD binding capacity.

The ExE-motif mutations also show interesting results. Previously it has been suggested that the hydrolysis of NAD and ADP-ribosylation of actin take place via an SN1 reaction, with Glu-387 playing a key role in oxocarbenium ion formation by interaction with the 2'OH ribose group of NAD. By creating E387Q, the –OH group of Glu-387 is substituted with $–NH_2$, which appears to completely abolish ADP-ribosylation of actin. This suggests that the proposed SN1 reaction involving Glu-387 has some validity. In addition E385Q also completely eliminates ADP-ribosylation of actin, which indicates that the –OH functional group plays a crucial role in the mechanism. The theory is that Glu-385 brings Arg-177 from actin into the active site via the –OH functional group of Glu-385. Again, this mutagenesis study indicates this theory may also be correct. Having already established that substitution of the functional group of Glu-387 to $–NH_2$, it confirms the view that increasing the size of this residue and substituting the functional group (E387R mutation) also eradicates ADP-ribosylation.

CDTa ADP-Ribosylation Mechanism Future Work

The purpose of mutating these specific three residues and monitoring their affect on cytotoxicity is to determine whether or not these residues are vital individually for the activity of CDTa. Here, the combination of both the Vero cell assays and the Western blot assay, have proven that altering these three residues can have a significant negative impact on the activity of CDTa. Therefore, these three residues should be key in small molecule drug design to block CDTa activity. If the activity of CDTa can be inhibited, then ADP-

ribosylation of actin can be prevented, which in turn will protect from cell rounding and cell death.

Solving the crystal structures of these CDTa mutants may shed some light on the orientation of the mutants in the active site, to ensure they are not having an adverse affect on cytotoxicity of CDTa based on their position, rather than the changes to their properties. At present the optimised protocol of CDTa native production and mutant production differs from that used for the original protocol that lead to successful determination of the native structure. Attempts have been made at crystallising the native and mutant proteins using the same conditions that were in place for the established protocol. However, as the protein is now expressed from a different construct there may be small changes to the terminal residues, and potentially changes to the overall fold of the proteins that may cause them to behave differently. Therefore, in order to grow diffraction quality crystals for structural studies, it may be necessary to screen for alternative crystallisation conditions for these constructs, which is currently in progress.

Chapter 5

Discussion and Future Work

5.1 Toxin A

Toxin A is a large multi-modular clostridial toxin with dominant hydrophobic properties. The four domains that make up Toxin A include the glucosylating 'A' domain, the binding 'B' domain, the cysteine protease 'C' domain and the hydrophobic `D' domain (Jank and Aktories, 2008). Each of these four domains plays a key role in toxicity (see Figure 1.2.1b in Chapter 1.2 for an illustration). The 'B' domain interacts with cell surface receptors on target cells and the whole toxin is taken up by receptor-mediated endocytosis. Within the endosomal compartment, pH changes stimulate conformational changes within Toxin A allowing pore formation of the 'D' domain and insertion into the endosomal membrane. The 'A' and 'C' domains are translocated into the cytosol, whereby the 'C' domain interacts with InsP6. The 'C' domain catalytically cleaves off the 'A' domain releasing it into the cytosol whereby it targets Rho GTPases for glucosylation (Davies *et al.*, 2011; Jank and Aktories, 2008). As discussed in Chapters 1.2 and 2.2, the crystal structures of the 'A' domain, 'C' domain and a fragment of the 'B domain for Toxin A have been solved at high resolution. The focal point of the Toxin A Ph.D project was to structurally characterise the full-length Toxin A using X-ray crystallography.

Although ambitious due to the size and nature of the toxin, attempts were made at crystallising Toxin A for use in X-ray crystallography studies. Toxin A contains a high percentage of leucine, isoleucine, phenylalanine and tyrosine residues, which gives the protein hydrophobic properties commonly seen in membrane proteins. The majority of the hydrophobic residues are located in the 'D' domain, as this is the membrane insertion domain. At present the 'D" domain has not been characterised structurally and similarly there are no structural homologues for this region. This hydrophobic element may be one of the reasons for the difficulties experienced during crystallisation. In addition, the 'B' domain structure reveals a helical region which is thought to increase surface area for optimized binding to target cells (Greco *et al.*, 2006). This region binds carbohydrate receptors on target cell surfaces. It is long, helical and may be quite flexible, which may explain why the protein molecules in the crystal are not packed tightly.

After extensive crystallisation trials and optimisations, diffraction quality crystals were obtained, and three datasets were collected, with the best resolution being 10 Å. Although ambitious, attempts were made at solving the structure by molecular replacement, however the combination of low-resolution data and lack of model for 50 % of the toxin is a challenging combination. There are potentially up to 12 molecules in the asymmetric unit, so if the resolution of the crystal diffraction can be improve to at least 5 Å or higher then it may be possible to solve the structure by molecular replacement.

Due to the difficulties experienced in high-resolution structural studies for Toxin A, low-resolution structural techniques were employed, including small-angle X-ray scattering (SAXS). SAXS data collected for Toxin A were used for *ab initio* modelling to produce a low-resolution structure, displayed in Figure 2.3a-A. The structure is consistent with the Toxin A reconstruction generated from electron microscopy data, Figure 2.3a-B (Pruitt *et al.*, 2010). The SAXS envelope of Toxin A was superimposed onto the model structure of Toxin A generated from electron microscopy data, Figure 2.3a-C, which clearly shows their similarities. Both structures have a globular head domain and an extended long tail domain. The SAXS structure of Toxin B, solved by Albesa-Jové *et al.*, shows few similarities to that of Toxin A other than that they are both elongated and approximately 265-275 Å. The Toxin B structure lacks the globular head domain and instead has a protruding domain mid way along the envelope, see Figure 2.2.1d (Albesa-Jové *et al.*, 2010). There has not been a successful 3D-reconstructed structure reported for Toxin B from electron microscopy data as yet. Previous attempts found that the electron microscopy images of Toxin B provided a more heterogeneous field of particles rather than the significantly more homogeneous field of particles reported for Toxin A (Pruitt *et al.*, 2010). The characteristic globular head was still present amongst the heterogeneous particles, but due to the heterogeneity of the sample, a model was not constructed.

Future Work for LCTs

There are numerous avenues in which this project could be taken further given more time. First and foremost, the continued work in crystallisation and X-

ray crystallography studies for Toxin A. The research performed during the project to date has included identification of a crystallisation condition for full-length Toxin A. Now that a condition has been established, further crystallisation condition optimisations should be carried out, including the use of alternative crystallisation techniques including microdialysis, crystallisation under oil, sandwich drop and free interface diffusion. The next stage in this progression should be to try co-crystallising Toxin A with various molecules to stabilise the protein. This could include carbohydrates or cell surface receptors for the 'B' domain or antibody fragments such as Fab fragments. In addition, complex studies of the 'A' domain and Rho GTPases could be performed, which could provide insights into the interaction between the glucosylation target and the 'A' domain. Highlighting these interactions using X-ray crystallography could provide a platform for specific drug design to inhibit glucosylation of Rho GTPases by Toxin A. Finally, as there is no crystallographic data for the full-length Toxin B, this structure should also be pursued given it is thought to be the most potent and toxic of these two LCTs.

Concluding Remarks

To conclude, elucidating the 3D-structure of Toxin A proved challenging using crystallographic techniques, however success was achieved using low-resolution structure techniques. A SAXS structure of Toxin A was obtained, which provided our first visualisation of Toxin A from X-ray scattering data. This structure was consistent with the model reconstruction structure obtained by electron microscopy. This achievement has brought this project one step further to elucidating the high-resolution full-length structure of Toxin A.

5.2 CDT Binary Toxin

Some more virulent *C.difficile* strains produce an ADP-ribosylating binary toxin; CDT, alongside the two potent exotoxins; Toxin A and B. The role of the binary toxin in pathogenicity is unclear, however some studies have shown that CDT causes the formation of microtubule-based protrusions on the cell surface of target cells (Schwan *et al.*, 2009). It is postulated that these protrusions promote adherence of *C.difficile* to the intestinal epithelial cells (Schwan *et al.*, 2009). As discussed in detail in the introductory chapter, the binary toxin is made up of two independently produced components: CDTa, the enzymatic component and CDTb, the transport component. The transport component is thought to interact with cell surface receptors on target cells. This component forms oligomers, interacts with the enzymatic component, CDTa, and both components are taken up into the cells by receptor mediated endocytosis. Studies have shown that individually CDTa and CDTb are non-toxic, however when the components are combined CDT is toxic to Vero cells (Sundriyal *et al.*, 2010).

The objectives for the binary toxin project were firstly to characterise the enzymatic mechanism of CDTa by testing the current hypothesis, secondly to solve the crystal structures of the CDTb variants and finally to form complexes of CDTa-CDTb and CDT-actin. Two of these objectives were achieved: the first being the mutagenesis studies performed on CDTa to elucidate the roles of active site residues in the ADP-ribosylation mechanism and the second being the formation of a CDTa-actin complex.

For the CDTa mechanism studies, see Chapter 4.2, both Western blot assays and the Vero cell assays were used to test the effect of active site mutations on the ribosylation mechanism. The results show that Ser-345, Glu-385 and Glu-387 all individually play essential roles in the mechanism. Based on the range of mutations performed, it was also observed that on removal of the –OH groups from both Glu-385 and Glu-387, ADP-ribosylation of actin was completely inhibited, which indicates that the proposed SN1 reaction involving Glu-385 and Glu-387 is correct.

Chapter 4.1 describes the steps taken to produce a CDT-actin complex. Although success was achieved in formation of the complex, the yield was too low to perform crystallisation trials. In addition, attempts made at forming a CDTa-CDTb complex were unsuccessful. The reasoning behind this lack of complex formation for CDTa-CDTb is unknown, however one contributing factor could be the discovery that CDTb variants form interchangeable oligomers of high molecular weight when concentrated. The way in which CDTa and CDTb interact is yet to be established. However, it is postulated that CDTb'' interacts with specific carbohydrate receptors on target cells and forms heptamers prior to CDTa binding (Barth, 2004; Barth *et al.*, 2004). It is unclear whether or not the conformation of CDTb changes upon receptor binding, or if receptor binding is required prior to interaction with CDTa.

Extensive crystallisation attempts were made for all three forms of CDTb and although diffraction quality crystals were grown for CDTb' and CDTb'', the resolution was low and the crystals did not survive in the beam long enough to collect data. Crystallisation occurs when proteins interact non-covalently to form highly ordered lattices. If proteins are irregular in shape the 'building blocks' of the crystal lattice do not pack well, which can lead to poor diffracting crystals. This might be the cause of the low-resolution diffraction observed for CDTb crystals. High concentrations of protein are required for crystallisation, however as discovered in Chapter 3.3, CDTb forms heterogeneous oligomers when concentrated. This may prevent tight and well-ordered crystal packing, which must be overcome in order to grow improved diffraction quality crystals.

Future Work for CDTa-actin Complex and CDTa Mechanism

The preliminary CDTa-actin complex formation has shown promising results for future work. The current 'on-column' protocol established in Chapter 4.1 should be pursued and performed on large scale, as this could potentially provide enough complex for crystallisation trials. If this complex can be crystallised and data collected, the structures of actin and CDTa individually can be used as models for molecular replacement. This structure could provide a platform for small molecular drug design to block ADP-ribosylation of actin. As it has been established in this project that the SN1 mechanism is used by CDTa to ADP-ribosylate actin, using Vero

cell and Western Blot assays, it would be useful to solve the crystal structures of these mutant CDTa constructs to supplement the biochemical assay data. This is required to confirm that the active site mutations are affecting the cytotoxicity of CDT based on their functional group alterations rather than the positioning of the mutated residues. High-resolution crystal structures of the mutants would confirm the position of these residues in the active site.

Future Work for CDTa-CDTb Complex and CDTb Variants

In terms of solving the structure of CDTb and forming a CDT-CDTb complex, future efforts should be focused on inhibition of the sporadic oligomerisation that occurs during concentration. This oligomerisation may prevent protein from forming tightly packed crystals, hence the lack of high-resolution diffraction from CDTb variants and may also prevent the interactions of CDTb'' with CDTa.

Concluding Remarks

Although the specific role of the binary toxin is still unknown, it is understood that the more virulent strains of *C.difficile* produce both of the LCTs and the binary toxin, which indicates that this toxin does contribute to pathogenicity. Studies have shown that the binary toxin can cause microtubule based protrusions on the target cell surfaces, increasing bacterial adherence to the gut, and that the binary toxin is toxic to Vero cells (Schwan *et al.*, 2009; Sundriyal *et al.*, 2010). The work in this thesis is directed towards a greater understanding of the actions of the binary toxin inside and outside of the target cells. Firstly investigating the method of CDTa entry into cells, including CDTb processing and oligomerisation and CDTa-CDTb interaction, and secondly the action of CDTa once in the cytosol in terms of mechanism of action and interactions with the ribosylation target. The achievements of this part of the project include the formation of a CDTa-actin complex which has provided a platform for future structural studies of this complex, in addition to generating evidence to conclude that the mechanism of ADP-ribosylation of actin by CDTa is an SN1 mechanism requiring residues Glu-385 and Glu-387 of the ExE-motif.

References

References

Albesa-Jové, D., Bertrand, T., Carpenter, E.P., Swain, G. V, Lim, J., Zhang, J., Haire, L.F., Vasisht, N., Braun, V., Lange, A., von Eichel-Streiber, C., Svergun, D.I., Fairweather, N.F., Brown, K.A., 2010. Four Distinct Structural Domains in Clostridium difficile Toxin B Visualized Using SAXS. *Journal of Molecular Biology* 396, 1260–1270.

Ananthakrishnan, A.N., 2011. Clostridium difficile Infection: Epidemiology, Risk Factors and Management. *Nature Reviews. Gastroenterology & Hepatology* 8, 17–26.

Arakawa, T., Timasheff, S., 1985. Theory of Protein Solubility. *Methods in Enzymology* 114, 49–77.

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics* 22, 195–201.

Aslam, S., Hamill, R., Musher, D., 2005. Treatment of Clostridium difficile-Associated Disease: Old Therapies and New Strategies. *The Lancet Infectious Diseases* 5, 549–557.

Aubry, A., Hussack, G., Chen, W., KuoLee, R., Twine, S.M., Fulton, K.M., Foote, S., Carrillo, C.D., Tanha, J., Logan, S.M., 2012. Modulation of Toxin Production by the Flagellar Regulon in Clostridium difficile. *Infection and Immunity* 80, 3521–3532.

Barth, H., 2004. Uptake of Binary Actin ADP-Ribosylating Toxins. *Reviews of Bhysiology, biochemistry and Pharmacology* 152, 165–182.

Barth, H., Aktories, K., Popoff, M.R., Stiles, B.G., 2004. Binary Bacterial Toxins: Biochemistry, Biology, and Applications of Common Clostridium and Bacillus Proteins. *Microbiology and Molecular Biology Reviews* 68, 373–402.

Barth, H., Pfeifer, G., Hofmann, F., Maier, E., Benz, R., Aktories, K., 2001. Low pH-Induced Formation of Ion Channels by Clostridium Difficile Toxin B in Target Cells. *The Journal of Biological Chemistry* 276, 10670–10676.

Bartlett, J.G., 2002. Clinical Practice. Antibiotic-Associated Diarrhea. *The New England Journal of Medicine* 346, 334–339.

Bergfors, T., 2009. Protein Crystallization, Second. ed. International University Line, La Jolla, California.

Bishop, A.L., Hall, A., 2000. Rho GTPases and Their Effector Proteins. *The Biochemical journal* 348 Pt 2, 241–255.

Blundell, T.L., Johnson, L.N., 1976. Protein Crystallography. Academic Press Inc.

Boistelle, R., 1986. The Concepts of Crystal Growth from Solution. *Advances in Nephrology from the Necker Hospital* 15, 173–217.

Boistelle, R., Astier, J.P., 1988. Crystallization Mechanisms in Solution. *Journal of Crystal Growth* 90, 14–30.

Borriello, S.P., 1998. Pathogenesis of Clostridium difficile Infection. *The Journal of Antimicrobial Chemotherapy* 41 Suppl C, 13–19.

Busch, C., Aktories, K., 2000. Microbial Toxins and the Glycosylation of Rho Family GTPases. *Current Opinion in Structural Biology* 10, 528–535.

Carroll, K., Bartlett, J., 2011. Biology of Clostridium difficile: Implications for Epidemiology and Diagnosis. *Annual Review of Microbiology* 65, 501–521.

Cecil, J.A., 2012. Clostridium difficile: Changing Epidemiology, Treatment and Infection Prevention Measures. *Current Infectious Disease Reports* 14, 612–619.

Ciesla, W.P., Bobak, D.A., 1998. Clostridium difficile Toxins A and B are Cation-Dependent UDP-Glucose Hydrolases with Differing Catalytic Activities. *The Journal of Biological Chemistryhemistry* 273, 16021–16026.

Compton, L.A., Johnson, W.C., 1986. Analysis of Protein Circular Dichroism Spectra for Secondary Structure Using a Simple Matrix Multiplication. *Analytical Biochemistry* 155, 155–167.

D'Urzo, N., Malito, E., Biancucci, M., Bottomley, M.J., Maione, D., Scarselli, M., Martinelli, M., 2012. The Structure of Clostridium difficile Toxin A Glucosyltransferase Domain Bound to Mn2+ and UDP Provides Insights into Glucosyltransferase Activity and Product Release. *The FEBS Journal* 279, 3085–3097.

Davies, A.H., Roberts, A.K., Shone, C.C., Acharya, K.R., 2011. Super Toxins from a Super Bug: Structure and Function of Clostridium difficile Toxins. *The Biochemical Journal* 436, 517–526.

Dove, C.H., Wang, S.Z., Price, S.B., Phelps, C.J., Lyerly, D.M., Wilkins, T.D., Johnson, J.L., 1990. Molecular Characterization of the Clostridium difficile Toxin A Gene. *Infection and Immunity* 58, 480–488.

Egerer, M., Giesemann, T., Jank, T., Satchell, K.J.F., Aktories, K., 2007. Auto-Catalytic Cleavage of Clostridium difficile Toxins A and B Depends on Cysteine Protease Activity. *The Journal of Biological Chemistryiological chemistry* 282, 25314–25321.

Evans, P., 2006. Scaling and assessment of data quality. *Acta Crystallographica. Section D, Biological crystallography* 62, 72–82.

Feigelson, R.S., 1988. The Relevance of Small Molecule Crystal Growth Theories and Techniques to the Growth of Biological Macromolecules. *Journal of Crystal Growth* 90, 1–13.

Franke, D., Svergun, D.I., 2009. DAMMIF, a Program for Rapid *Ab-initio* Shape Determination in Small-Angle Scattering. *Journal of Applied Crystallography* 42, 342–346.

Frisch, C., Gerhard, R., Aktories, K., Hofmann, F., Just, I., 2003. The Complete Receptor-Binding Domain of Clostridium difficile Toxin A is Required for Endocytosis. *Biochemical and Biophysical Research Communications* 300, 706–711.

Geric, B., Johnson, S., Gerding, D.N., Grabnar, M., Rupnik, M., 2003. Frequency of Binary Toxin Genes Among Clostridium difficile Strains That Do Not Produce Large Clostridial Toxins. *Journal of Clinical Microbiology* 41, 5227–5232.

Giesemann, T., Egerer, M., Jank, T., Aktories, K., 2008. Processing of Clostridium difficile Toxins. *Journal of Medical Microbiology* 57, 690–696.

Glatter, O., 1977. A New Method for the Evaluation of Small-Angle Scattering Data. *Journal of Applied Crystallography* 10, 415–421.

Govind, R., Dupuy, B., 2012. Secretion of Clostridium difficile Toxins A and B Requires the Holin-like Protein TcdE. *PLoS Pathogens* 8, e1002727.

Greco, A., Ho, J.G.S., Lin, S.-J., Palcic, M.M., Rupnik, M., Ng, K.K.-S., 2006. Carbohydrate Recognition by Clostridium difficile Toxin A. *Nature Structural & Molecular Biology* 13, 460–461.

Gulke, I., Pfeifer, G., Liese, J., Fritz, M., Hofmann, F., Aktories, K., Barth, H., 2001. Characterization of the Enzymatic Component of the ADP-Ribosyltransferase Toxin CDTa from Clostridium difficile. *Infection and Immunity* 69, 6004–6011.

Hall, I.C., O'Toole, E., 1935. Intestinal Flora in New-Born Infants with a Description of a New Pathogenic Anaerobe Bacillus difficilis. *The American Jounral of Diseases in Children* 49, 390–402.

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D.J., Martin, M.J., Connor, T.R., Harris, S.R., Fairley, D., Bamford, K.B., D'Arc, S., Brazier, J., Brown, D., Coia, J.E., Douce, G., Gerding, D., Kim, H.J., Koh, T.H., Kato, H., Senoh, M., Louie, T., Michell, S., Butt, E., Peacock, S.J., Brown, N.M., Riley, T., Songer, G., Wilcox, M., Pirmohamed, M., Kuijper, E., Hawkey, P., Wren, B.W., Dougan, G., Parkhill, J., Lawley, T.D., 2013. Emergence and Global Spread of Epidemic Healthcare-Associated Clostridium difficile. *Nature Genetics* 45, 109–113.

Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., Minton, N.P., 2007. The ClosTron: a Universal Gene Knock-Out System for the Genus Clostridium. *Journal of Microbiological Methods* 70, 452–464.

Ho, J.G.S., Greco, A., Rupnik, M., Ng, K.K.-S., 2005. Crystal Structure of Receptor-Binding C-Terminal Repeats from Clostridium difficile Toxin A. *Proceedings of the National Academy of Sciences of the United States of America* 102, 18373–18378.

Holbourn, K.P., Shone, C.C., Acharya, K.R., 2006. A Family of Killer Toxins. Exploring the Mechanism of ADP-Ribosylating Toxins. *The FEBS Journal* 273, 4579–4593.

Jacques, D.A., Trewhella, J., 2010. Small-Angle Scattering for Structural Biology--Expanding the Frontier While Avoiding the Pitfalls. *Protein Science : a Publication of the Protein Society* 19, 642–657.

Jank, T., Aktories, K., 2008. Structure and Mode of Action of Clostridial Glucosylating Toxins: the ABCD Model. *Trends in Microbiology* 16, 222–229.

Jank, T., Giesemann, T., Aktories, K., 2007a. Rho-Glucosylating Clostridium difficile Toxins A and B: New Insights into Structure and Function. *Glycobiology* 17, 15R–22R.

Jank, T., Giesemann, T., Aktories, K., 2007b. Clostridium difficile Glucosyltransferase Toxin B-Essential Amino Acids for Substrate Binding. *The Journal of Biological Chemistry* 282, 35222–35231.

Jank, T., Reinert, D.J., Giesemann, T., Schulz, G.E., Aktories, K., 2005. Change of the Donor Substrate Specificity of Clostridium difficile Toxin B by Site-Directed Mutagenesis. *The Journal of Biological Chemistry* 280, 37833–37838.

Just, I., Gerhard, R., 2004. Large Clostridial Cytotoxins. *Reviews of Physiology, Biochemistry and Pharmacology* 152, 23–47.

Just, I., Seizer, J., Wilm, M., von Eichel-Streiber, C., Mann, M., Aktories, K., 1995a. Glucosylation of Rho Proteins by Clostridium difficile Toxin B. *Letters to Nature* 375, 500–503.

Just, I., Wilm, M., Selzer, J., Rex, G., von Eichel-Streiber, C., Mann, M., Aktories, K., 1995b. The Enterotoxin from Clostridium difficile (ToxA) Monoglucosylates the Rho Proteins. *The Journal of Biological Chemistry* 270, 13932–13936.

Kabsch, W., 2010. XDS. *Acta Crystallographica. Section D, Biological crystallography* 66, 125–132.

Kelly, C., LaMont, T., 1998. Clostridium difficile Infection. *Annual Review of Medicine* 49, 375–390.

Kim, J., Pai, H., Seo, M.-R., Kang, J.O., 2012. Clinical and Microbiologic Characteristics of tcdA-Negative Variant Clostridium difficile Infections. *BMC Infectious Diseases* 12, 109.

Kintzer, A.F., Thoren, K.L., Sterling, H.J., Dong, K.C., Feld, G.K., Tang, I.I., Zhang, T.T., Williams, E.R., Berger, J.M., Krantz, B. a, 2009. The Protective Antigen Component of Anthrax Toxin Forms Functional Octameric Complexes. *Journal of Molecular Biology* 392, 614–629.

Konarev, P., Volkov, V., Sokolova, A., Kock, M., Svergun, D.I., 2003. PRIMUS: a Windows PC-Based System for Small-Angle Scattering Data Analysis. *Journal of Applied Crystallography* 36, 1277–1282.

Kreimeyer, I., Euler, F., Marckscheffel, A., Tatge, H., Pich, A., Olling, A., Schwarz, J., Just, I., Gerhard, R., 2011. Autoproteolytic Cleavage Mediates Cytotoxicity of Clostridium difficile Toxin A. *Naunyn-Schmiedeberg's Archives of Pharmacology* 383, 253–262.

Krivan, H.C., Clark, G.F., Smith, D.F., Wilkins, T.D., 1986. Cell Surface Binding Site for Clostridium difficile Enterotoxin : Evidence for a Glycoconjugate Containing the Sequence. *Infection and Immunity* 53, 573–581.

Kuehne, S.A., Cartman, S.T., Heap, J.T., Kelly, M.L., Cockayne, A., Minton, N.P., 2010. The Role of Toxin A and Toxin B in Clostridium difficile Infection. *Nature* 467, 711–713.

Kuehne, S.A., Cartman, S.T., Minton, N.P., 2011. Both, Toxin A and Toxin B, Are Important in Clostridium difficile Infection. *Gut Microbes* 2, 252–255.

Kyne, L., 2010. Clostridium difficile: Beyond Antibiotics. *New England Journal of Medicine* 362, 264–265.

Lorber, B., Fischer, F., Bailly, M., Roy, H., Kern, D., 2012. Protein Analysis by Dynamic Light Scattering : Methods and Techniques for Students. *Biochemistry and Molecular Biology Education* 40, 372–382.

Lyerly, D.M., Krivan, H.C., Wilkins, T.D., 1988. Clostridium difficile : Its Disease and Toxins. *Clinical Microbiology Reviews* 1, 1–18.

Lyerly, D.M., Saum, K.E., MacDonald, D.K., Wilkins, T.D., 1985. Effects of Clostridium difficile Toxins Given Intragastrically to Animals. *Infection and Immunity* 47, 349–352.

Lyras, D., O'Connor, J.R., Howarth, P.M., Sambol, S.P., Carter, G.P., Phumoonna, T., Poon, R., Adams, V., Vedantam, G., Johnson, S., Gerding, D.N., Rood, J.I., 2009. Toxin B is Essential for Virulence of Clostridium difficile. *Nature* 458, 1176–1179.

Matamouros, S., England, P., Dupuy, B., 2007. Clostridium difficile Toxin Expression is Inhibited by the Novel Regulator TcdC. *Molecular microbiology* 64, 1274–1288.

Mattila, E., Uusitalo-Seppälä, R., Wuorela, M., Lehtola, L., Nurmi, H., Ristikankare, M., Moilanen, V., Salminen, K., Seppälä, M., Mattila, P.S., Anttila, V.-J., Arkkila, P., 2012. Fecal Transplantation, Through Colonoscopy, is Effective Therapy for Recurrent Clostridium difficile Infection. *Gastroenterology* 142, 490–496.

Mauss, S., Chaponnier, C., Just, I., Aktories, K., Gabbiani, G., 1990. ADP-Ribosylation of Actin Isoforms by Clostridium botulinum C2 Toxin and Clostridium perfringens Iota Toxin. *European Journal of Biochemistry / FEBS* 194, 237–241.

McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L., Read, R., 2007. Phaser Crystallographic Software. *Journal of Applied Crystallography* 40, 658–674.

McGlone, S., Bailey, R., Zimmer, S., Popovich, M., Tian, Y., Ufberg, P., Muder, R., Lee, S., 2012. The Economic Burden of Clostridium difficile. *Clinical Microbiology and Infection* 18, 282–289.

McMaster-Baxter, N.L., Musher, D.M., 2007. Clostridium difficile: Recent Epidemiologic Findings and Advances in Therapy. *Pharmacotherapy* 27, 1029–1039.

McPherson, A., 1989. Science in Pictures: Macromolecular Crystals. *Scientific American* 260, 62–69.

Murshudov, G., Vagin, A., Dodson, E., Read, R., 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica. Section D, Biological crystallography* D53, 240–255.

Musher, D., Aslam, S., Logan, N., Nallacheru, S., Bhaila, I., Borchert, F., Hamill, R., 2005. Relatively Poor Outcome After Treatment of Clostridium difficile Colitis with Metronidazole. *Clinical Infectious Diseases* 40, 1586–1590.

Nagahama, M., Sakaguchi, Y., Kobayashi, K., Ochi, S., Sakurai, J., 2000. Characterization of the Enzymatic Component of Clostridium perfringens Iota-Toxin. *Journal of Bacteriology* 182, 2096–2103.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *Journal of molecular biology* 302, 205–17.

O'Connor, J.R., Lyras, D., Farrow, K.A., Adams, V., Powell, D.R., Hinds, J., Cheung, J.K., Rood, J.I., 2006. Construction and Analysis of Chromosomal Clostridium difficile Mutants. *Molecular Microbiology* 61, 1335–1351.

Papatheodorou, P., Carette, J.E., Bell, G.W., Schwan, C., Guttenberg, G., Brummelkamp, T.R., Aktories, K., 2011. Lipolysis-Stimulated Lipoprotein Receptor (LSR) is the Host Receptor for the Binary Toxin Clostridium difficile Transferase (CDT). *Proceedings of the National Academy of Sciences of the United States of America* 108, 16422–16427.

Pellizzari, R., Rossetto, O., Schiavo, G., Montecucco, C., 1999. Tetanus and botulinum Neurotoxins: Mechanism of Action and Therapeutic Uses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 354, 259–268.

Perelle, S., Domenighini, M., Popotp, M.R., 1996. Evidence that Arg-295 , Glu-378 , and Glu-380 are Active-Site Residues of the ADP-Ribosyltransferase Activity of Iota Toxin. *FEBS letters* 395, 191–194.

Perelle, S., Gibert, M., Bourlioux, P., Corthier, G., Popoff, M.R., 1997. Production of a Complete Binary Toxin (Actin-Specific ADP-Ribosyltransferase) by Clostridium difficile CD196. *Infection and Immunity* 65, 1402–1407.

Popoff, M.R., Rubin, E.J., Gill, D.M., Boquet, P., 1988. Actin-Specific ADP-Ribosyltransferase Produced by a Clostridium difficile Strain. *Infection and Immunity* 56, 2299–2306.

Pruitt, R.N., Chagot, B., Cover, M., Chazin, W.J., Spiller, B., Lacy, D.B., 2009. Structure-Function Analysis of Inositol Hexakisphosphate-induced Autoprocessing in Clostridium difficile Toxin A. *The Journal of Biological Chemistry* 284, 21934–21940.

Pruitt, R.N., Chambers, M.G., Ng, K.K., Ohi, M.D., Lacy, D.B., 2010. Structural Organization of the Functional Domains of Clostridium difficile Toxins A and B. *Proceedings of the National Academy of Sciences of the United States of America* 107, 13467–13472.

Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A., 2007. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Quarterly Reviews of Biophysics* 40, 191–285.

Qa'Dan, M., Spyres, L.M., Ballard, J.D., 2000. pH-Induced Conformational Changes in Clostridium difficile Toxin B. *Infection and Immunity* 68, 2470–2474.

Reinert, D.J., Jank, T., Aktories, K., Schulz, G.E., 2005. Structural Basis for the Function of Clostridium difficile Toxin B. *Journal of Molecular Biology* 351, 973–981.

Rupnik, M., 2005. Revised Nomenclature of Clostridium difficile Toxins and Associated Genes. *Journal of Medical Microbiology* 54, 113–117.

Rupnik, M., Avesani, R., Janc, M., von Eichel-Streiber, C., Delmee, M., 1998. A Novel Toxinotyping Scheme and Correlation of Toxinotypes with Serogroups of Clostridium difficile Isolates. *Journal of Clinical Microbiology* 36, 2240–2247.

Rupnik, Maja, Pabst, S., Rupnik, Marjan, von Eichel-Streiber, C., Urlaub, H., Söling, H.D., 2005. Characterization of the Cleavage Site and Function of Resulting Cleavage Fragments After Limited Proteolysis of Clostridium difficile Toxin B (TcdB) by Host Cells. *Microbiology* 151, 199–208.

Schleberger, C., Hochmann, H., Barth, H., Aktories, K., Schulz, G.E., 2006. Structure and Action of the Binary C2 Toxin from Clostridium botulinum. *Journal of Molecular Biology* 364, 705–715.

Schwan, C., Stecher, B., Tzivelekidis, T., Ham, M. Van, Rohde, M., Hardt, W.D., Wehland, J., Aktories, K., 2009. Clostridium difficile Toxin CDT Induces Formation of Microtubule-Based Protrusions and Increases Adherence of Bacteria. *PLoS pathogens* 5, 1–14.

Shen, A., 2012. Clostridium difficile Toxins: Mediators of Inflammation. *Journal of Innate Immunity* 4, 149–158.

Shen, A., Lupardus, P.J., Puri, A.W., Albrow, V.E., Malte, M., Garcia, K.C., Bogyo, M., 2011. Defining an Allosteric Circuit in the Cysteine Protease Domain of Clostridium difficile Toxins. *Nature Stucture and Molecular Biology* 18, 364–371.

Sundriyal, A., Roberts, A.K., Ling, R., McGlashan, J., Shone, C.C., Acharya, K.R., 2010. Expression, Purification and Cell Cytotoxicity of Actin-Modifying Binary Toxin from Clostridium difficile. *Protein Expression and Purification* 74, 42–48.

Sundriyal, A., Roberts, A.K., Shone, C.C., Acharya, K.R., 2009. Structural Basis for Substrate Recognition in the Enzymatic Component of ADP-Ribosyltransferase Toxin CDTa from Clostridium difficile. *The Journal of Biological Chemistry* 284, 28713–28719.

Svergun, D., 1992. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* 25, 495–503.

Tan, K.S., Wee, B.Y., Song, K.P., 2001. Evidence for Holin Function of tcdE Gene in the Pathogenicity of Clostridium difficile. *Journal of Clinical Microbiology* 50, 613–619.

The PyMOL Molecular Graphics System, version 1.5.0.4,

Tsuge, H., Nagahama, M., Nishimura, H., Hisatsune, J., Sakaguchi, Y., Itogawa, Y., Katunuma, N., Sakurai, J., 2003. Crystal Structure and Site-directed Mutagenesis of Enzymatic Components from Clostridium perfringens Iota-toxin. *Journal of Molecular Biology* 325, 471–483.

Tsuge, H., Nagahama, M., Oda, M., Iwamoto, S., Utsunomiya, H., Marquez, V.E., Katunuma, N., Nishizawa, M., Sakurai, J., 2008. Structural Basis of Actin Recognition and Arginine ADP-Ribosylation by Clostridium perfringens Iota-Toxin. *Proceedings of the National Academy of Sciences of the United States of America* 105, 7399–7404.

Ünligil, U.M., Rini, J.M., 2000. Glycosyltransferase Structure and Mechanism. *Current Opinion in Structural Biology* 10, 510–517.

Vagin, A., Teplyakov, A., 1997. MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography* 30, 1022–1025.

Van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., van Grondelle, R., Groen, F.C., 1990. Estimation of Protein Secondary Structure and Error Analysis From Circular Dichroism Spectra. *Analytical Biochemistry* 191, 110–118.

Vandekerckhove, J., Schering, B., Barmann, M., Aktories, K., 1988. Botulinum C2 Toxin ADP-ribosylates Cytoplasmic β/Ɣ-Actin in Arginine 177. *The Journal of Biological Chemistry* 263, 696–700.

Vandekerckhove, J., Schering, B., Bärmann, M., Aktories, K., 1987. Clostridium perfringens Iota Toxin ADP-Ribosylates Skeletal Muscle Actin in Arg-177. *FEBS letters* 225, 48–52.

Vedantam, G., Clark, A., Chu, M., Mcquade, R., Mallozzi, M., Viswanathan, V.K., 2012. Clostridium difficile Infection: Toxins and Non-Toxin Virulence Factors, and Their Contributions to Disease Establishment and Host Response. *Gut Microbes* 3, 121–134.

Volkov, V., Svergun, D.I., 2003. Uniqueness of Ab-initio Shape Determination in Small-Angle Scattering. *Journal of Applied Crystallography* 36, 860–864.

Voth, D.E., Ballard, J.D., 2005. Clostridium difficile Toxins : Mechanism of Action and Role in Disease. *Clinical Microbiology Reviews* 18, 247–263.

Wang, D., Chen, H., Barber, P., Wyatt, P., 1979. Light Scattering by Polydisperse Suspensions of Inhomogenous Nonspherical Particles. *Applied Optics* 18, 2672–2678.

Winter, G., 2010. xia2: An Expert System for Macromolecular Crystallography Data Reduction. *Journal of Applied Crystallography* 43, 186–190.

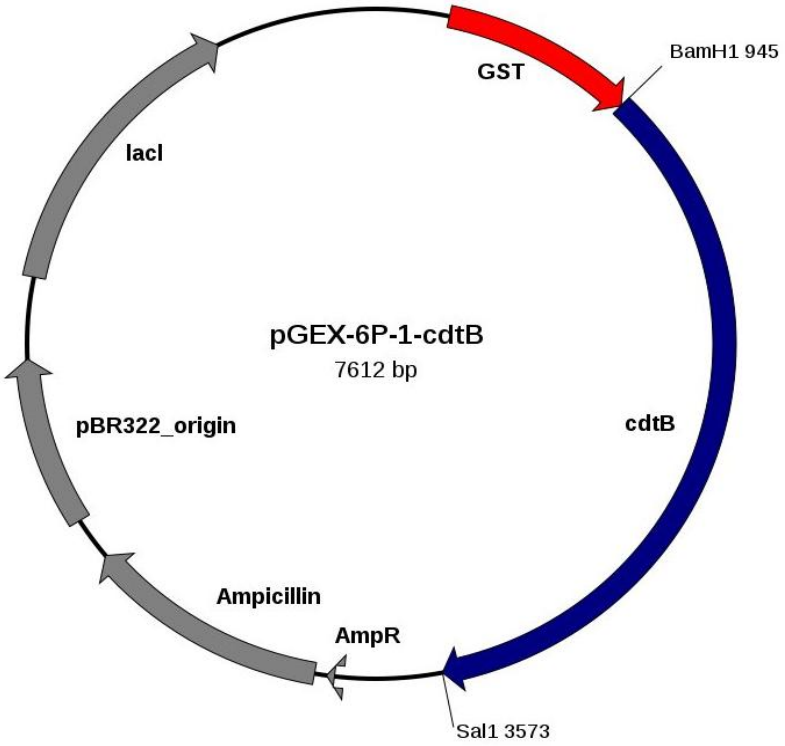
Wiscombe, W., 1980. Improved Mie Scattering Algorithms. *Applied Optics* 19, 1505–1509.

Wyatt, P., 1998. Submicrometer Particle Sizing by Multiangle Light Scattering Following Fractionation. *Journal of Colloid and Interface Science* 197, 9–20.

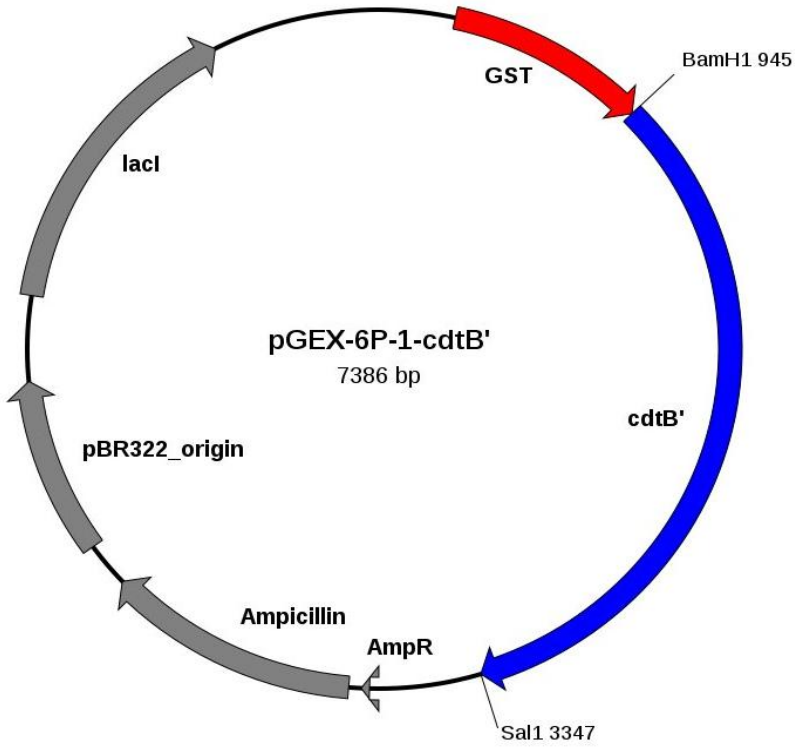
Appendices

Appendix 1- Plasmid Maps

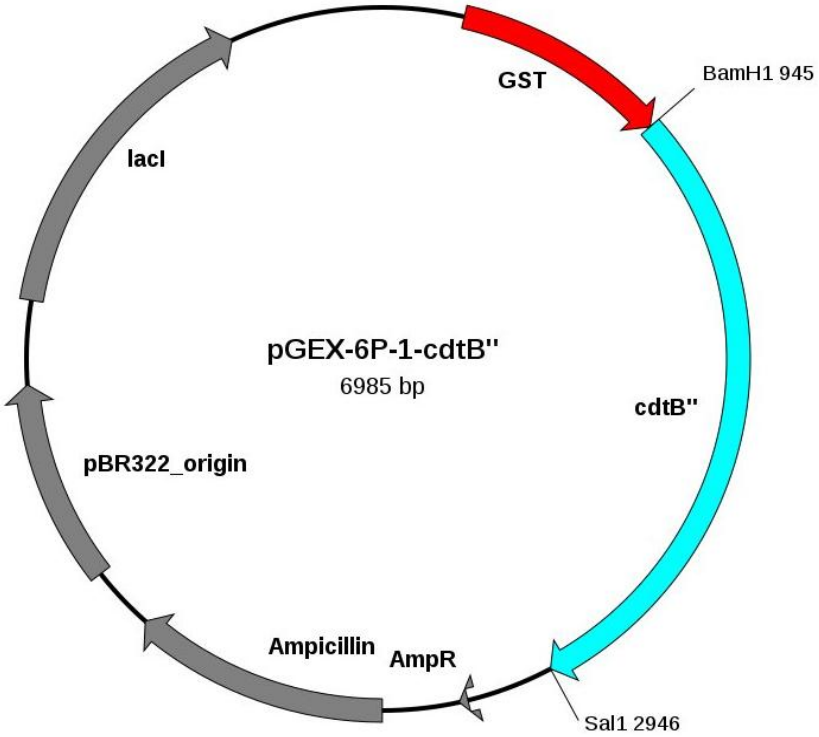
CDTb



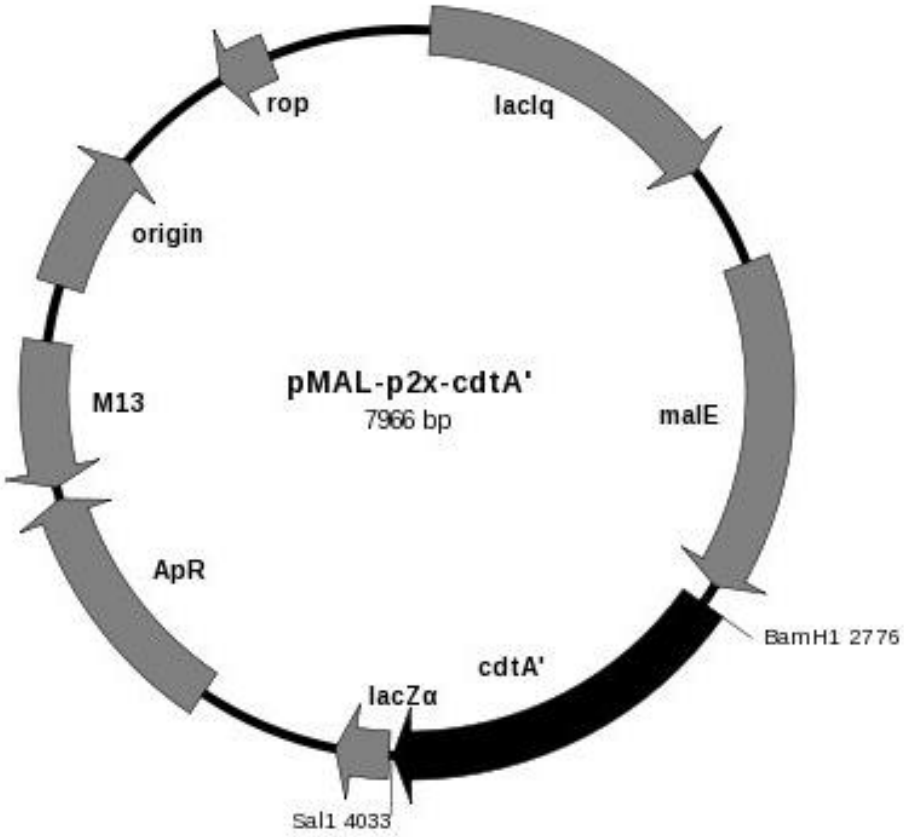
CDTb’



CDTb''


GST
BamH1 945
lacI
pGEX-6P-1-cdtB''
6985 bp
cdtB''
pBR322_origin
Ampicillin
AmpR
Sal1 2946

CDTa'


rop
laclq
origin
M13
pMAL-p2x-cdtA'
7966 bp
malE
ApR
BamH1 2776
cdtA'
lacZα
Sal1 4033

Appendix 2- Recipes for Media

Lysogeny Broth

For 1 L of LB medium:

10 g tryptone

5 g yeast extract

10 g sodium chloride

Terrific Broth

For 1 L of TB medium:

12 g tryptone

24 g yeast extract

4 ml glycerol

(make up to 900 ml and add 100 ml of TB salts)

For 100 ml TB salts:

2.13 g potassium phosphate monobasic

12.54 g potassium phosphate dibasic

Appendix 3- Sequencing of CDTa Mutants

```
Native    338- LSYPNFISTSIGSVNMSAFA- 357
S345A     338- LSYPNFIATSIGSVNMSAFA- 357
                ******* ******* * * ****

Native    338- LSYPNFISTSIGSVNMSAFA- 357
S345F     338- LSYPNFIFTSIGSVNMSAFA- 357
                ******* ******* * *****


Native    338- LSYPNFISTSIGSVNMSAFA- 357
S345R     338- LSYPNFIRTSIGSVNMSAFA- 357
                ******* ******* * *****

Native    338- LSYPNFISTSIGSVNMSAFA- 357
S345Y     338- LSYPNFIYTSIGSVNMSAFA- 357
                ******* ******** *****
```

```
Native    378- AIPGYAGEYEVLLNHGSKFK- 397
E385Q     378- AIPGYAGQYEVLLNHGSKFK- 397
                *** *** * : ******* *****

Native    378- AIPGYAGEYEVLLNHGSKFK- 397
E387Q     378- AIPGYAGEYQVLLNHGSKFK- 397
                *** *** ** *: ***** *****

Native    378- AIPGYAGEYEVLLNHGSKFK- 397
E387R     378- AIPGYAGEYRVLLNHGSKFK- 397
                *** ****** . **********
```

BJ www.biochemj.org

BJ Signal

Biochem. J. (2011) **436**, 517–526 (Printed in Great Britain) doi:10.1042/BJ20110106

multimedia

REVIEW ARTICLE

Super toxins from a super bug: structure and function of *Clostridium difficile* toxins

Abigail H. DAVIES*†, April K. ROBERTS†, Clifford C. SHONE† and K. Ravi ACHARYA*[1]

*Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, U.K., and †Health Protection Agency, Porton Down, Salisbury SP4 0JG, U.K.

Clostridium difficile, a highly infectious bacterium, is the leading cause of antibiotic-associated pseudomembranous colitis. In 2009, the number of death certificates mentioning *C. difficile* infection in the U.K. was estimated at 3933 with 44 % of certificates recording infection as the underlying cause of death. A number of virulence factors facilitate its pathogenicity, among which are two potent exotoxins; Toxins A and B. Both are large monoglucosyltransferases that catalyse the glucosylation, and hence inactivation, of Rho-GTPases (small regulatory proteins of the eukaryote actin cell cytoskeleton), leading to disorganization of the cytoskeleton and cell death. The roles of Toxins A and B in the context of *C. difficile* infection is unknown. In addition to these exotoxins, some strains of *C. difficile* produce an unrelated ADP-ribosylating binary toxin. This toxin consists of two independently produced components: an enzymatic component (CDTa) and the other, the transport component (CDTb) which facilitates translocation of CDTa into target cells. CDTa irreversibly ADP-ribosylates G-actin in target cells, which disrupts the F-actin:G-actin equilibrium leading to cell rounding and cell death. In the present review we provide a summary of the current structural understanding of these toxins and discuss how it may be used to identify potential targets for specific drug design.

Key words: large clostridial toxin (LCT), Toxin A (TcdA), Toxin B (TcdB), ADP-ribosyltransferase (ADPRT), *Clostridium difficile* binary toxin (CDT), pseudomembranous colitis (PMC).

INTRODUCTION

Clostridium difficile is an anaerobic Gram-positive spore-forming bacterium that was first described in 1935. It was given this name as it grew slowly in culture and was difficult to isolate [1]. *C. difficile* was not thought to be a particularly harmful pathogen until a rise in the number of cases of PMC (pseudomembranous colitis) in the 1970s. Infection with *C. difficile* accounts for 90–100 % of antibiotic-associated PMC, with a 6–30 % mortality rate [2]. The majority of patients diagnosed with PMC are treated with broad-spectrum antibiotics such as metronidazole and vancomycin, which can be very effective. Metronidazole is the drug of choice as it is a lower cost drug than vancomycin. Metronidazole and vancomycin have similar efficacy in mild disease, however there is growing concern regarding treatment failure with the use of metronidazole in severe disease. Vancomycin has been shown to be superior in treatment of severe disease [3]. The use of broad-spectrum antibiotics to treat other diseases can disturb the balance of the normal bacterial flora in the gut, resulting in host susceptibility to colonization or overgrowth of *C. difficile* [1,4–6]. It is estimated that over 50 % of infants are colonized with *C. difficile*. However, these infants are usually asymptomatic and do not develop PMC. At present there is no clear understanding of why the infection does not manifest in infants [7]. It is thought that the infants' intestinal cells are not fully developed and so are deficient in toxin-specific carbohydrate surface receptors [5,8]. Patients infected with *C. difficile* initially experience mild diarrhoea, abdominal pains and a fever, and are diagnosed by the presence of *C. difficile* toxins in stool samples. In the case of mild infection, patients are usually advised to discontinue taking any antibiotics to prevent imbalance of flora in the gut. However, in extreme cases of *C. difficile* infection leading to PMC, metronidazole and vancomycin are the suggested treatments. Relapse is a common problem associated with antibiotic treatment (occurring in up to 25 % of cases), in which case patients are required to continue with prolonged courses of these drugs [4]. *C. difficile* is a highly infectious bacterium, and its spores can survive on surfaces for long periods of times, therefore it is a difficult organism to control in hospital environments and so the majority of infections are contracted nosocomially.

VIRULENCE FACTORS

C. difficile has a number of virulence factors that contribute to its pathogenicity, but this review will focus on its secreted toxins. The pathogenic strains of *C. difficile* produce two potent exotoxins, Toxin A and Toxin B (often called TcdA and TcdB), which induce mucosal inflammation and diarrhoea [1,6,9]. In addition to these exotoxins, some *C. difficile* strains produce an ADP-ribosylating binary toxin (CDT), however, the role of this toxin in disease is unclear [10,11]. In the present review, we focus on the current structural knowledge of clostridial toxins and relevant data regarding the roles of these structures within the host will be discussed in detail.

LARGE CLOSTRIDIAL TOXINS (LCTs)

Biology of LCTs

Bacteria from the genus *Clostridium* produce a wide range of toxins, the most well known being the clostridial neurotoxins such

Abbreviations used: CD-grease, α-Gal-(1,3)-β-Gal-(1,4)-β-GlcNAcO$(CH_2)_8CO_2CH_3$; CDT, *Clostridium difficile* ADP-ribosylating binary toxin; CPD, cysteine protease 'C' domain; GAP, GTPase-activating protein; GDI, guanine-nucleotide-dissociation inhibitor; GEF, guanine-nucleotide-exchange factor; LCT, large clostridial toxin; PaLoc, pathogenicity locus; PMC, pseudomembranous colitis; SAXS, small-angle X-ray scattering; TNS, 2-(*p*-toluidinyl) napthalene-6-sulfonic acid, sodium salt; UDP-Glc, UDP-glucose.

[1] To whom correspondence should be addressed (email k.r.acharya@bath.ac.uk).

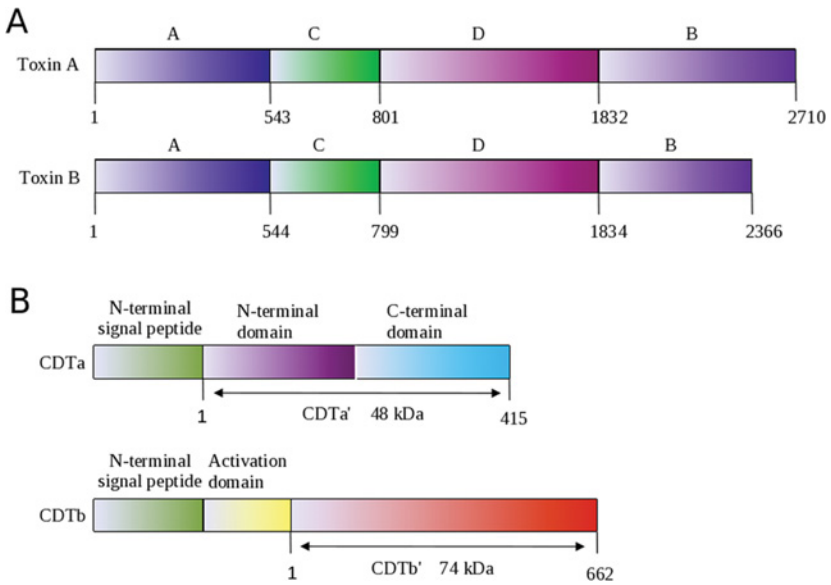


Figure 1 Domain organization of *C. difficile* toxins

(**A**) The LCTs are composed of four domains, the glucosylating enzymatic domain 'A' (dark blue), the autocatalytic processing domain 'C' (green), the translocating domain 'D' (magenta) and the binding domain 'B' (purple). The current solved crystal structures include the 'A' domain of Toxin B (Figure 3), the 'C' domain of Toxin A (Figure 5) and the 'B' domain of Toxin A (Figure 4). The colour scheme for the domain organization of the LCTs is maintained throughout all of the Figures. (**B**) The domain organization of CDT is shown for the two independently produced components, CDTa and CDTb. The N-terminal signal peptides are displayed in green for both components. The active CDTa' domain is coloured in accordance with Figure 6, with the N-terminal domain (purple) and the C-terminal domain (cyan). The CDTb component has an activation domain (yellow) that must be cleaved to give the CDTb' activated domain (red).

as the BoNTs (*Clostridium botulinum* neurotoxins) which can cause flaccid muscular paralysis known as botulism, and the TeNT (*Clostridium tetani* neurotoxin) which causes spastic paralysis [12]. Toxins A and B, however, are part of the LCT group, owing to their high molecular mass: Toxin A is 308 kDa and Toxin B is 269.6 kDa [13]. All of the members of the LCT family, which includes Toxins A and B from *C. difficile*, lethal toxins from *Clostridium sordellii* and α-toxin from *Clostridium novyi*, act on target cells by modifying small GTPases [14,15]. Toxins A and B are encoded by the genes *tcdA* and *tcdB*, which are located in a 19.6 kb locus, known as PaLoc (pathogenicity locus). There are three other genes in the PaLoc that encode positive and negative regulators of the production of these toxins, in addition to a gene encoding a holin-like protein which allows secretion of the toxins [16,17]. *C. difficile* strains are grouped on the basis of the variations in the structure of the PaLoc [16,18]. There are a number of reports that suggest Toxin B is essential for pathogenicity, whereas the role of Toxin A is less clear. This is reflected in the reported outbreaks of *C. difficile* strains that are Toxin A negative/Toxin B positive. Universal gene knock-out systems developed for the genus *Clostridium* [19,20] were used in studies that indicate Toxin A is not essential for virulence [21,22]. However, a recent report suggests that Toxin A is able to cause disease in the absence of Toxin B [23].

Toxins A and B have a multi-modular domain structure described by Jank and Aktories [13] as the ABCD model. This comprises: (A) the biologically active N-terminal domain; (B) the C-terminal binding domain; (C) the cysteine protease domain; and (D) the hydrophobic domain (for a schematic diagram see Figure 1A). The C-terminal binding 'B' domain, which consists of polypeptide repeats, is involved in receptor binding to specific cell-surface carbohydrate receptors, but other domains may also play a direct or indirect role in binding [24]. The precise mechanism of toxin uptake into the cells is unclear; however, there is evidence to suggest that it is a four step procedure, summarized in Figure 2: (1) highly specific binding of the 'B' domain to receptors on the cell surface followed by receptor-mediated endocytosis of the toxin into the endosomal compartment; (2) a decrease in pH within the endosomal compartment causes conformational changes within the LCTs, allowing pore formation and the subsequent translocation of part of the toxin into the cytosol, which has been shown by the use of ^{86}Rb ions [25,26]. The 'D' domain is a large hydrophobic region that makes up almost 50 % of the total size of the toxin and is thought to be involved in pore formation prior to translocation of the 'A' domain into the cytosol [25,26]. (3) The toxins undergo autoproteolysis allowing only the enzymatic 'A' domain to be released into the cytosol. It is thought that the cysteine protease domain 'C' is involved in this autoprocessing, and requires InsP_6 [27–29]; and finally (4) Rho proteins are targeted for glucosylation by the biologically active 'A' domain in the cytosol. Toxins A and B target Rho GTPases (Rho, Ras and Cdc42), which are molecular switches involved in numerous signal processes, in particular, the regulation of the actin cytoskeleton (Figure 2). Once the toxins enter the cytosol, they catalyse the addition of UDP-Glc (UDP-glucose) to Thr^{37} (monoglucosylation) in Rho GTPase leading to depolymerization of actin filaments, disruption of the cytoskeleton and eventually cell rounding and cell death [30–32]. *C. difficile* is a harmful pathogen and the main cause of the rise in cases of PMC in hospitals, therefore it is of great interest to understand the function of LCTs. Structural studies have given us an insight into the molecular mechanism of the LCTs, and we will discuss these below.

Structural aspects of LCTs

Negative stain electron microscopy

Recent studies have shown that Toxins A and B (which share 68 % sequence similarity and 47 % sequence identity) both have similar native structures to each other using negative stain electron microscopy [33]. Toxins A and B both have an elongated

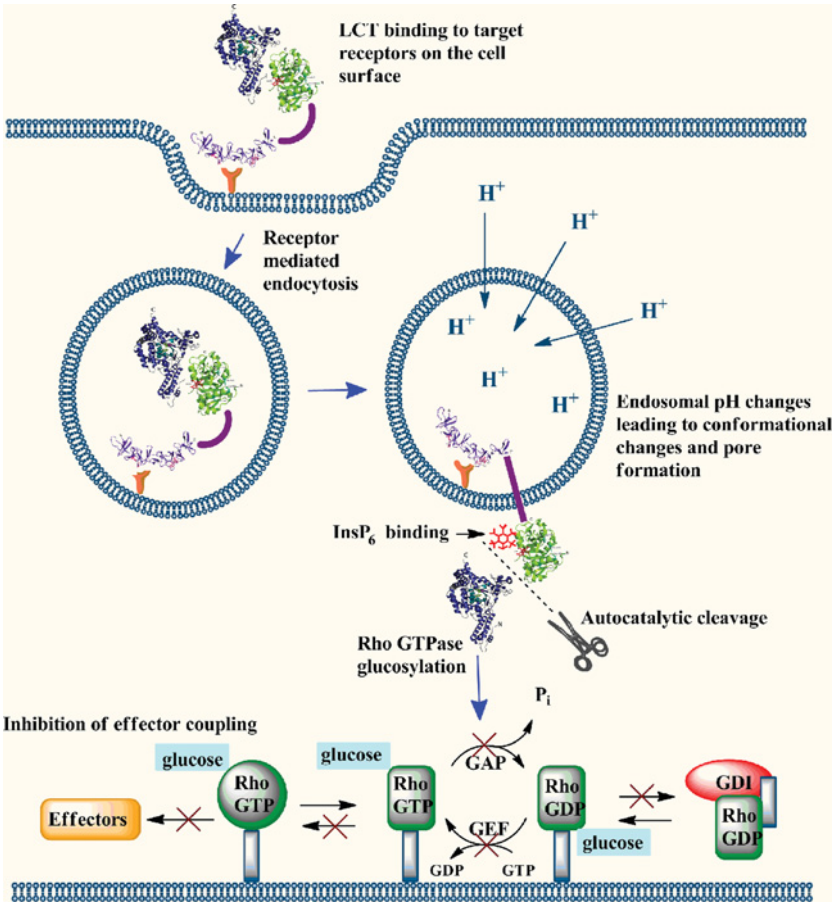


Figure 2 Cartoon representation of the processing machinery of Clostridial toxins

The LCT is displayed using the three-dimensional structures known for the four domains (note: the D domain structure is unknown) from both Toxins A and B, and is colour co-ordinated to the primary domain structure shown in Figure 1. The toxins bind to the target cells via the B domain (purple) and are taken up by receptor-mediated endocytosis. Following a change in the endosomal pH, conformational changes occur within the toxin and the D domain (magenta) forms a pore in the endosomal membrane allowing translocation of the A domain (dark blue) into the cytosol. This then undergoes autocatalytic cleavage via the C domain (green) on binding of InsP_6. Upon autocatalytic cleavage, the A domain is released into the cytosol where it glucosylates target Rho GTPases. In the GDP-bound conformations, Rho GTPases are inactive. GDIs keep GTPases in the cytosol. Rho GTPases are activated by GEFs, enabling interaction with effectors that control a number of signalling pathways. GAPs facilitate the hydrolysis of GTP to GDP, converting GTPases back to their inactive form. The A domain (dark blue) glucosylates Rho GTPases, which prevents effector coupling thereby inhibiting specific signalling pathways. In addition, it blocks GEF and GAP activity. GTPases are no longer able to bind GDI and so are found at the plasma membrane. An animated version of this Figure is available at http://www.BiochemJ.org/bj/436/0517/bj4360517add.htm.

structure containing a 'head' domain, a long 'tail' domain and a short inner 'tail' domain. As Toxin A was considerably more homogenous in structure than Toxin B, mapping studies have focused on Toxin A. The three-dimensional structure of Toxin A was constructed using the random conical tilt approach where the modular domains were mapped into the corresponding structure [33]. After imaging the fragments of the individual domains using electron microscopy and antibody labelling, Pruitt et al. [33] concluded that the long 'tail' fragment corresponds to the 'B' domain, the short 'tail' corresponds to the 'A' domain and the 'head' corresponds to the 'D' domain. The cysteine protease domain was not located. The same structure was then reviewed using a lower pH environment and, as predicted, conformational changes were observed in the 'head' shape, confirming the location of the 'D' domain in the structure, and also in the glucosyltransferase 'A' domain, providing a framework for the molecular mechanism of translocation [33]. Although negative stain electron microscopy provides images of the overall toxin shape, it is beneficial to understand the toxins at a molecular level using X-ray crystallography. Owing to the size and nature of the LCTs, purification and crystallization of the whole toxins is difficult and so a significant amount of work by researchers is focused on structure determination of the four individual domains.

Small angle X-ray scattering

Recently reported data using SAXS (small-angle X-ray scattering) have provided an *ab initio,* low resolution surface model of the native Toxin B [34]. When combined with the known structural information obtained and models of the unknown domains, this information has allowed visualization of the actual organization of the four individual domains of LCTs for the first time. The data revealed four distinct domain boundaries, into which the structural information and model structures can be aligned. The four domains are organised as demonstrated in Figure 1(A). The hydrophobic 'D' domain contains a highly solvent-exposed region which protrudes away from the core structure [34]. The SAXS data were collected at a neutral pH, so the low pH-induced conformational change that is predicted to occur within this 'D' domain has not been shown in this structure.

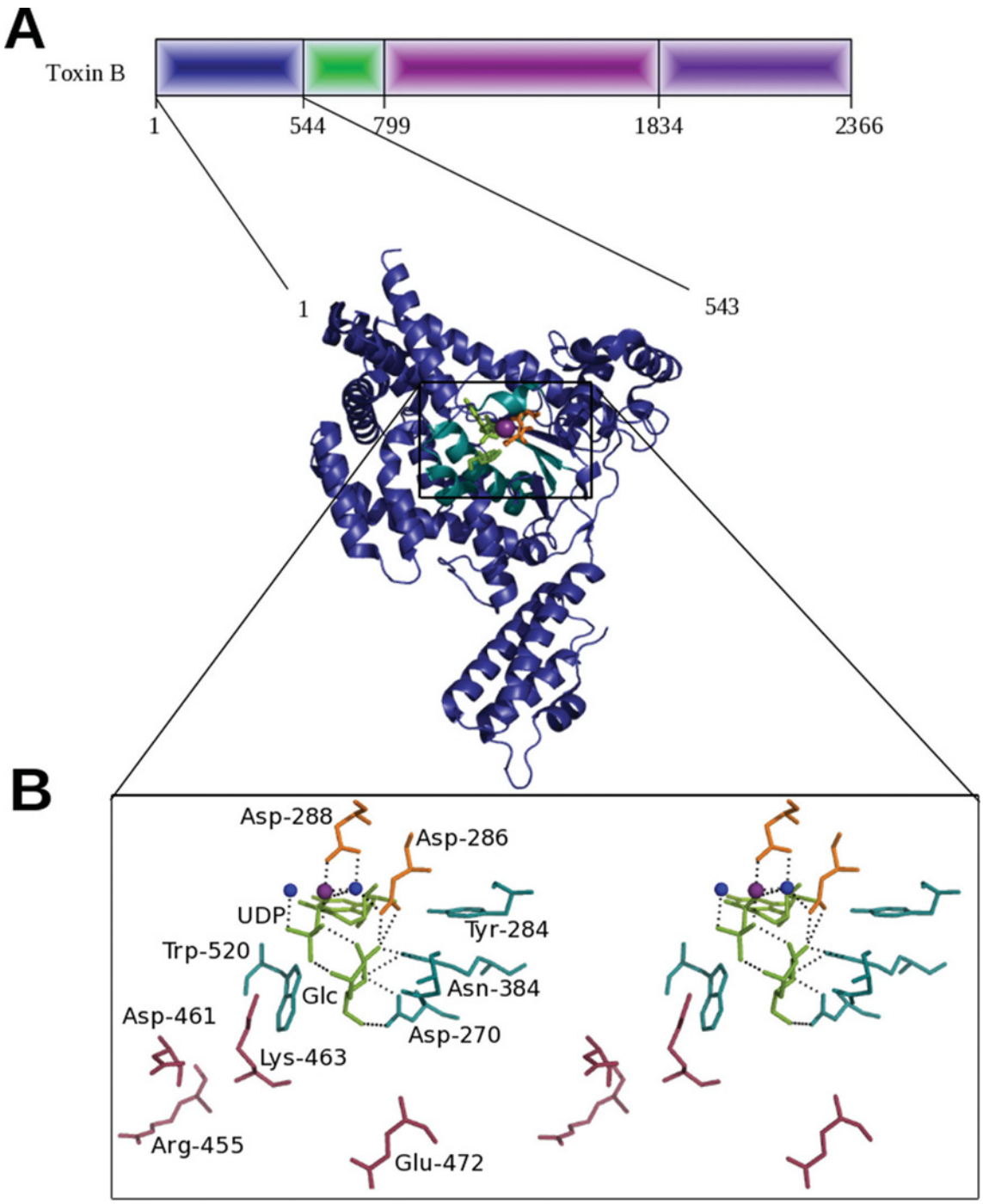


Figure 3 The glucosyltransferase 'A' domain of Toxin B

(**A**) Crystal structure of the glucosyltransferase domain of *C. difficile* Toxin B (dark blue), where UDP-Glc (green) is bound to the catalytic cleft via a manganese ion (magenta), with the residues involved in the active site displayed (teal) (PDB code 2BVL) [34]. (**B**) Stereo view of the catalytic cleft of Toxin B with water molecules (blue) and manganese ion (magenta). Residues involved in enzyme activity (teal) and UDP-Glc recognition (raspberry) are displayed, in addition to the DxD motif (orange). Images were created using PyMOL (http://www.pymol.org).

X-ray crystallographic studies of individual LCT domains

'A' domain: glucosyltransferase N-terminal domain

The glucosyltransferase domain, which is responsible for glucosylation of Rho proteins, is located at the N-terminal region known as the 'A' domain. As there is limited structural data for both Toxins A and B, examples have been taken from the individual domains, where available, for analysis in the present review. Figure 3(A) displays the crystal structure of the 'A' domain of Toxin B which gives some insights into the mechanistic implications of Toxin B activity within target cells (PDB code 2BVL) [35]. The structure of the 'A' domain of Toxin A is yet to be determined. This catalytic fragment of the LCT is delivered to the cytosol where it glucosylates small GTPases. The 'A' domain of Toxin B consists of residues 1–543 in the VPI 10463 *C. difficile* strain and was co-crystallized with UDP-Glc and

manganese ions that are essential for catalysis [36,37]. UDP-Glc was cleaved during the crystallization process owing to the hydrolytic activity of the toxin, but the products α-D-glucose and UDP were identified in the electron density [35]. The crystal structure shares the same common fold (consisting of 234 residues) seen among the glucosyltransferase type A family. The remaining 309 residues are arranged predominantly in α-helices and are thought to contribute to the specificity of the toxin [35]. As shown in Figure 3(B), the catalytic pocket in this Toxin B domain is made up of the β2, β5 and β10 strands, in addition to α12 and α18 helices, and a 510–523 residue loop [38]. The DXD motif involved in the binding of UDP-Glc and Mn^{2+} is composed of Asp^{286} and Asp^{288}. The Asp^{288} residue binds directly to Mn^{2+}, whereas Asp^{286} forms a hydrogen bond with a water molecule of the Mn^{2+} co-ordination sphere, in addition to binding to both the 3′-hydroxy group of UDP-ribose and glucose, making it a key residue for catalysis [35]. Mutational analysis by alanine scanning revealed five important residues involved in enzyme activity: Asp^{270}, Arg^{273}, Tyr^{284}, Asn^{384} and Trp^{520}. In addition, four residues, Arg^{455}, Asp^{461}, Lys^{463} and Glu^{472}, were found to be important for UDP-Glc recognition [39].

The LCTs target and glucosylate the Rho family of GTPase proteins. These are molecular switches that are predominantly responsible for regulating the actin cytoskeleton. These Rho proteins play other roles in cellular functions such as cell polarity, gene transcription and cell cycle progression. The Rho GTPase family of proteins consists of the Rho, Rac and Cdc42 subfamilies. The proteins exist in two forms: the GTP-bound form which is active and the GDP-bound form which is inactive, as displayed in Figure 2. These two forms are inter-convertible and are regulated by GEFs (guanine-nucleotide-exchange factors), GDIs (guanine-nucleotide-dissociation inhibitors) and GAPs (GTPase-activating proteins) (Figure 2). GDIs bind to Rho GTPases at a switch I region, rendering them soluble in the cytosol and in the inactive form. Activation of GTPases to their GTP-bound form is induced by GEFs, and this activation allows interaction with numerous effector molecules that control a number of signalling pathways. GAPs hydrolyse GTP-bound Rho GTPases into inactive GDP-bound Rho GTPases [40]. In the GDP-bound conformation, Rho GTPases have an exposed Thr^{37} residue which is targeted by LCTs for glucosylation. Upon glucosylation, actin filaments depolymerize, leading to disruption of the cytoskeleton and cell death. The Rho proteins are not glucosylated when bound to GTP or when their switch I region is complexed with GDIs, as the Thr^{37} residue is not exposed. Rho is glucosylated in the GDP-bound conformation as the Thr^{37} residue is exposed and allows an oxygen acceptor atom to be in position for attack from the C1 donor atom of the LCTs [31,32]. Toxin B binds to RhoA in the Rho-GTP form [41].

Two possible mechanisms have been suggested for the catalytic activity of the 'A' domain; a double displacement reaction (S_N2) and an internal return stereo-specific (S_N1-like) reaction. The double displacement reaction firstly requires a nucleophile in a position to attack the β-side of the C1 atom in glucose. However, as this atom is surrounded by non-polar residues, there is no nucleophile in position to attack the C1 atom, hence a double displacement reaction is unlikely. An internal return mechanism S_N1 is possible, as the oxocarbenium intermediate would be stabilized by the surrounding negative phosphate and carboxylate groups from Asp^{270}, Asp^{286}, Asp^{288} and Glu^{515}, which are in turn compensated by the positively charged Arg^{273} and a Mn^{2+} ion [35]. LCTs exhibit high substrate/co-substrate specificity, with Toxins A and B only binding UDP-Glc. The two residues Ile^{383} and Glu^{385}, which are located in the active site, are responsible for this specificity as shown by mutagenesis studies [39,41].

'B' domain: binding C-terminal domain

The initial step in the toxication process of *C. difficile* is the binding of LCTs to the cell-surface of intestinal epithelial cells, which is carried out by the 'B' domain. The 'B' domain of LCTs interacts with carbohydrate structures such as Gal-α-(1,3)-Gal-β-(1,4)-GlcNAc on the host epithelial cells [42]. Although a functional carbohydrate receptor in humans has yet to be identified, attempts to gain more insights into carbohydrate binding have been made. The crystal structure of the C-terminal binding domain of Toxin A has been determined for two different fragment sizes. The first fragment (127 residues), named TcdA-f1, from strain 48489, toxinotype VI was obtained by expressing a longer fragment and isolating a smaller cleaved fragment (PDB code 2F6E) [43]. The second fragment, TcdA-f2, which is a slightly longer fragment (255 residues), is in complex with a synthetic derivative of a carbohydrate receptor, CD-grease [α-Gal-(1,3)-β-Gal-(1,4)-β-$GlcNAcO(CH_2)_8CO_2CH_3$] (Figure 4A) (PDB code 2G7C) [44]. In this complex there are two carbohydrate-binding regions. However, in the full-length C-terminal fragment there are seven of these potential binding domains that are highly conserved, giving it a high binding capacity. Although there is little information about the binding domain of Toxin B, it is believed that Toxin B uses different receptors to bind to target cell surfaces than Toxin A [41]. However, considering Toxins A and B share 47 % sequence identity, the structural information regarding the Toxin A–carbohydrate interactions may provide some clues as to how Toxin B might interact with carbohydrate surface receptors.

Sequence analysis of the Toxin A gene shows that 31.5 % of the gene is in 38 contiguous repeating units and these repeats are located in the C-terminal region. These repeats can be split into two categories: class I (long), of which there are seven types, each repeat being 30 amino acids long; and class II (short), of which there are 31 types, each consisting of either 20 or 21 amino acids [45]. TcdA-f2 contains nine small repeats and two long repeats, whereas TcdA-f1 consists of four short repeats and one long repeat [43,44]. Each short and long repeat forms β-hairpins that are connected by loops of 7–10 residues in short repeats and 18 residues in long repeats [41]. The β-hairpins are related to each other by a 3_1 screw-axis transformation, meaning that each β-hairpin is related to the adjacent hairpin by 120° (Figure 4B) [44]. This results in a β-solenoid left-handed helix, which is a common conformation of many bacterial cell-surface-binding proteins, and is thought to increase surface area for optimized binding to the target cells (Figure 4B) [44]. These results are consistent with that of the $TcdAC_{26–39}$ structure of the 'B' domain from Albesa-Jové et al. [34].

'C' domain: cysteine protease domain and the 'D' domain: hydrophobic region

The CPD (cysteine protease 'C' domain) is thought to be located between residues 543 and 769 in Toxin A and between residues 543 and 767 in Toxin B [27]. The crystal structure of the CPD (543–809) from Toxin A was determined at 1.6 Å (1 Å = 0.1 nm) in the presence of $InsP_6$ (Figure 5A) (PDB code 3HO6) [29]. This sits between the enzymatic domain and the delivery domain, playing a role in proteolytic cleavage of the toxin. Toxins A and B both undergo auto-catalytic cleavage in the presence of $InsP_6$. Once the target cell has taken up the LCT via receptor-mediated endocytosis at the 'B' domain, the toxin undergoes autoproteolysis in order to allow the 'A' domain to pass across the endosomal membrane into the cytosol [46]. The structure reveals a central nine-stranded β-sheet flanked by five α-helices. The

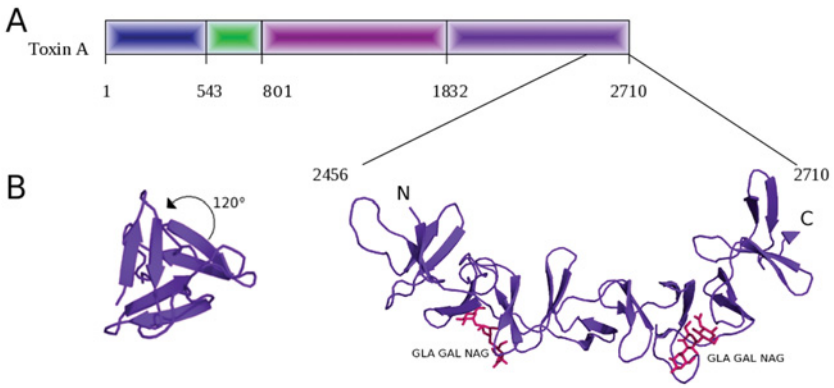


Figure 4 The binding 'B' domain of Toxin A

(**A**) Ribbon structure of the binding domain of *C. difficile* Toxin A (purple) with two carbohydrate synthetic derivatives of CD-grease bound, shown as stick models (pink) (PDB code 2G7C) [43]. (**B**) Cross-section through the binding domain of Toxin B showing the β-hairpin 3_1 screw axis, where the β-hairpins are related to the adjacent hairpin by 120°. Images were created using PyMOL (http://www.pymol.org). Adapted from Jank et al. [30] with permission © 2007 Oxford University Press.

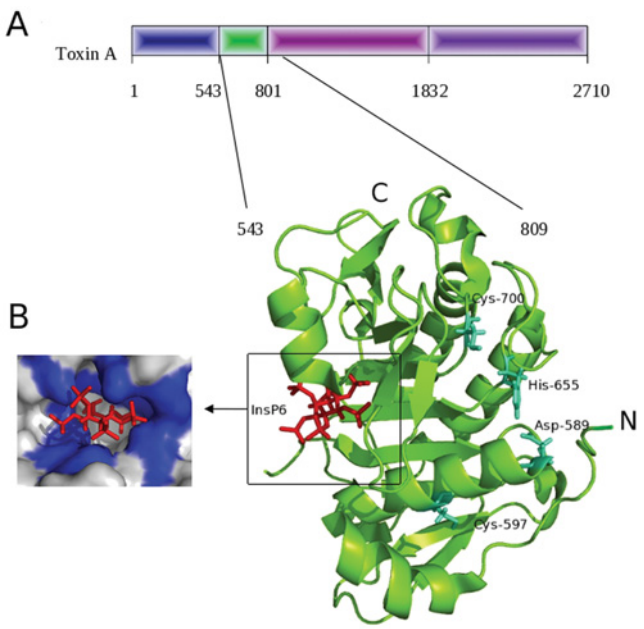


Figure 5 The cysteine protease 'C' domain of *C. difficile* Toxin A

(**A**) Ribbon structure of the 'C' domain (green) displaying the proposed residues in the active site (cyan) and the InsP_6 (red) bound on the opposite side of the central nine-stranded β-sheet (PDB code 3HO6) [28]. (**B**) The InsP_6-binding site on the 'C' domain displaying the electrostatic potential of the binding site (blue represents positively charged residues). InsP_6 is displayed as a ball and stick molecule (red). The InsP_6-binding site is highly positive as shown by the abundance of blue colour. Images were created using PyMOL (http://www.pymol.org).

InsP_6-binding site and the proposed active site are located on opposite sides of the central β-sheet and are separated by a three-stranded β-hairpin, known as the 'β-flap'. The N-terminal domain extends around the exterior of the domain from the InsP_6-binding site to the proposed active site [29]. Cys^{700}, His^{655} and Asp^{589} have been identified as the catalytic triad. However, owing to the distances between these residues, hydrogen bond formation is not possible as with other cysteine proteases. Therefore the mechanism of catalysis may differ from other cysteine proteases. Pruitt et al. [29] have shown that the CPD changed conformation on binding of InsP_6 to a more stable form, which increases resistance to chymotrypsin digestion. InsP_6 is a highly negatively charged ligand and binds to a number of positively charged residues that span the entire domain including Arg^{753}, Tyr^{579} and seven lysine residues: Lys^{577}, Lys^{602}, Lys^{649}, Lys^{754}, Lys^{766}, Lys^{777} and Lys^{794} (see Figure 5B, with these positively charged residues in blue). The 'β-flap' is thought to be involved in transmitting structural changes from the InsP_6-binding site across to the active site. The CPD uses InsP_6 to activate an intramolecular autoproteolytic cleavage event which allows the correct processing of the toxins to transfer the enzymatic domain into the cytosol. Shen et al. [46a] recently solved the structure of the CPD of Toxin B (PDB code 3PEE) which shares 56 % sequence identity with the CPD of Toxin A and a highly similar structure [rmsd (root mean square deviation) of 1.3 Å]. Comparison of the structures show a completely conserved active site, an almost identical InsP_6-binding site (with the exception of two residues) and a well-conserved β-flap. Using chemical probes, Shen et al. [46a] showed that the CPD of Toxin B can adopt the activated conformation without the presence of InsP_6, and InsP_6 in fact shifts the equilibrium in favour of the active conformation. Through the use of tryptophan fluorescence assays, mutagenesis and structural analysis. Shen et al. [46a] were able to show that a conserved group of interconnected residues in the β-flap region are responsible for communicating the InsP_6-binding signal to the active site.

Following receptor-mediated endocytosis of the LCTs, the 'A' domain is translocated into the cytosol. The hydrophobic 'D' domain is thought to be involved in the translocation, although little is known about its structure. LCTs undergo conformational changes stimulated by a decrease in endosomal pH. Qa'Dan et al. [26] used bafilomycin A1, which is a potent inhibitor of the endosomal vacuolar ATPase pump that controls the acidity of the endosome, to assess the effect of endosomal acidification. Translocation was measured using TNS [2-(*p*-toluidinyl) napthalene-6-sulfonic acid, sodium salt] fluorescence. For Toxin B, as the acidity of the endosome increased, an increase in TNS fluorescence was seen, suggesting a conformational change within Toxin B and exposure of the hydrophobic 'D' domain, thereby confirming its role in translocation [26]. Owing to the hydrophobic nature of the 'D' domain, purification and crystallization are difficult, which may explain why the structure has not yet been determined.

CLOSTRIDIUM DIFFICILE BINARY TOXIN

Biology of the binary toxin

In addition to the LCTs, some *C. difficile* strains can produce an additional toxin called the ADP-ribosyltransferase binary toxin, CDT, which is a member of the ADPRT family [11]. The ADPRT family is divided into four classes on the basis of domain organization and their targets [47]. The first group is the AB_5 group that includes cholera toxin. Within this group, the toxins are composed of one A subunit and five B subunits, and they target small regulatory G-proteins. The second group is the AB three domain group which includes diphtheria toxin. Toxins of this group ribosylate a diphthamide residue on elongation factor 2, and have a binding domain, a transmembrane domain and a catalytic domain. The third group is the single polypeptide group, which includes C3 toxin from *C. botulinum*, and have unknown roles in pathogenesis. The final group is the AB binary toxin group, which are made up of two subunits and target actin. This group includes the C2 toxin from *C. botulinum* and Iota toxin from *Clostridium perfringens* [47]. CDT is a member of the AB binary toxin group and is made up of two independently produced components: the enzymatic component, CDTa (mature length 48 kDa), and the transport component, CDTb (mature length 74 kDa); see Figure 1(B) for a schematic representation of the domain organization [10,48].

In one study of CDT genes in the United States, 15.5 % of *C. difficile* strains tested had both the genes for the CDT components (*cdtB* and *cdtA*), 8.7 % of which did not have the LCT genes [49]. The precise pathogenesis of CDT is yet to be established due to the dominant production of LCTs in most virulent strains of *C. difficile* and lack of CDT [10]. However, it has been shown that a CDTa–CDTb (CDT) complex is toxic to Vero cells [50]. A recent study has suggested a role for CDT in increasing adherence of bacteria to target cells, by the formation of microtubule protrusions. Using time-lapse and immunofluorescence microscopy, Schwan et al. [51], have shown that CDT forms dynamic microtubule protrusions on the surface of human colon carcinoma cells (Caco-2) concomitantly with ADP-ribosylation of actin and depolymerization of microfilaments. In addition to protrusion formation, cellular microtubule structures were also altered to increase bundling of microtubules. The CDT-induced formation of protrusions caused a ∼5-fold increase in adherence of *C. difficile* in anaerobic conditions. The protrusions were shown to form a dense mesh-work in which the bacteria were caught, contributing to the colonization of *C. difficile*. Similar results were also demonstrated for the homologues *C. botulinum* toxin C2 and *C. perfringens* Iota toxin [51].

The transport component, CDTb, is essential for entry of CDTa (enzymatic component) into the cytosol [52]. The proposed mechanism of uptake of CDTa is similar to other members of the ADPRT family such as *C. botulinum* C2 toxin and Iota toxin from *C. perfringens* [48]. The CDTb component must be activated via cleavage, after which it then can form heptamers at the cell surface and bind to specific cell-surface receptors. Subsequently, CDTa binds to CDTb and is taken up into the cell by receptor-mediated endocytosis. In a similar mechanism to the LCTs, low pH-induced conformational changes occur leading to potential heptameric pore formation and translocation of CDTa into the cytosol [48,52]. The N-terminus of CDTa is responsible for interaction with CDTb, whereas the C-terminus harbours the enzymatic activity (Figure 1B) [53]. We can predict that CDTa will irreversibly ADP-ribosylate monomeric G-actin at the Arg^{177} residue from the structural evidence of Iota toxin (closest homologue of CDTa). This ADP-ribosylation will block polymerization of G-actin to F-actin and subsequently disrupt the F-actin:G-actin equilibrium [54,55]. Disruption of the actin equilibrium results in cell rounding and cell death.

Structural aspects of CDT

The crystal structure of CDTa was solved at three different pH values, 4.0, 8.5 and 9.0 (PDB codes 2WN8, 2WN4 and 2WN5 respectively) [56]. In addition, the structure was solved in complex with ADP ribose donors NADPH and NAD at pH 9.0 (PDB codes 2WN6 and 2WN7) [56]. The structure in complex with NAD was determined at 2.25 Å and detailed mechanistic implications of CDTa have been proposed on the basis of this structural data [56]. It is known that CDTa transfers the ADP ribose group of NAD/NADPH to monomeric G-actin at Arg^{177}, blocking polymerization of actin and therefore leading to the collapse of the cell cytoskeleton. This prediction was made on the basis of the mechanism of *C. perfringens* Iota toxin, of which CDTa shares 84 % sequence identity [54,55].

The structure of CDTa in complex with NAD is displayed in Figure 6(A). We can observe the N-terminal domain, which consists of five α-helices and eight β-strands, extending from residues 1 to 215, and the C-terminal domain, which extends from residues 224 to 240, also consisting of five α-helices and eight β-strands. These domains are linked by a loop extending from residues 216 to 223, shown in red in Figure 6(A) [56]. The C-terminal domain harbours the enzymatic activity, whereas the N-terminal domain is predicted to interact with the CDTb domain. Both NAD and NADPH bind to the catalytic cleft of CDTa via the interacting residues Arg^{302}, Arg^{303}, Gln^{307}, Asn^{342} and Ser^{345} [56]. Figure 6(B) displays these catalytic residues in addition to a number of other defining features of the ADPRT family of toxins, for example the PN loop (orange) and the Arg-motif (yellow). The highly flexible ARTT loop can be seen in blue, which is an ADP-ribosyl turn-turn loop that is important for substrate binding, along with the catalytic residues of the EXE motif, which has been shown to be crucial for activity in ADPRTs. In CDTa, the EXE motif is composed of the residues Glu^{385} and Glu^{387}, which is thought to be involved in stabilizing the substrate–enzyme complex as seen with the corresponding residues Glu^{378} and Glu^{380} in Iota toxin. However, in this structure both of these residues are not in direct contact with NAD or NADPH, which might suggest that the EXE motif in CDTa is not necessary for ligand binding and stabilization of this complex, although there is no experimental evidence to support this [56]. However, this could be due to the location of both Glu^{385} and Glu^{387} on the flexible ARTT loop. The final feature shown in Figure 6(B) is the STS motif, which includes the residues Ser^{345} and Ser^{347}. Ser^{345} forms a strong hydrogen bond with Glu^{387}, and also directly with NAD and NADPH, therefore suggesting a role for the STS motif in ligand binding and catalysis [56].

The proposed mechanism of ADP-ribosylation of actin in Iota toxin was suggested to be an S_N1 reaction after formation of an actin–Iota complex and also by the use of site-directed mutagenesis [57]. The S_N1 reaction occurs via two intermediates: first an oxocarbenium ion intermediate and secondly a cationic intermediate. This mechanism is thought to be a common mechanism amongst the ADPRT family of toxins [47]. As Iota toxin is the closest homologue to CDTa, a similar S_N1 reaction has been proposed [56]. In following the S_N1 mechanism, the catalytic glutamate Glu^{387} in CDTa should form a H-bond with the 2′-OH of ribose forming an oxocarbenium intermediate and rendering the ribose group vulnerable to nucleophilic attack (Figure 7). Following cleavage of NAD and formation of an oxocarbenium ion, it is thought the ARTT loop is rearranged to bring the

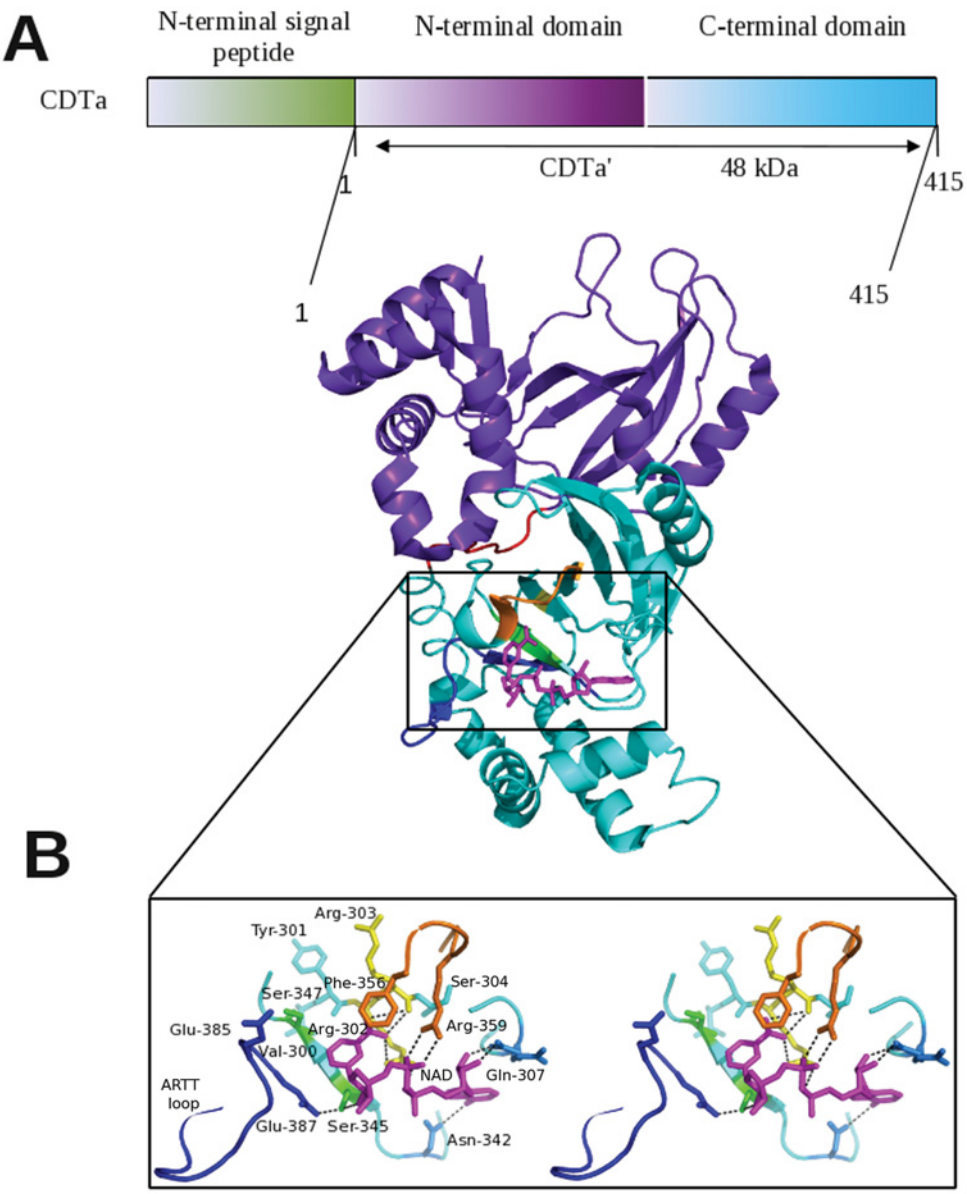


Figure 6 The active CDTa component of CDT from *C. difficile*

(**A**) Ribbon structure displaying the N-terminal domain (purple), the C-terminal domain (cyan) and the connecting loop between the C-terminal and N-terminal domains (red), and the active site is shown within the black box (PDB code 2WN7) [55]. (**B**) Stereo view of the active site of CDTa with key residues involved in NAD binding: STS-motif (green), the PN-loop (orange), the EXE motif and ARTT loop (blue), the Arg-Arg residues (yellow), and any remaining residues involved in binding are shown in light blue and the NAD molecule is in pink. Images were created using PyMOL (http://www.pymol.org).

Glu^{385} residue to the reaction centre to stabilize the transfer of ADP-ribose to the Arg^{177} of actin [56]. Further structural and experimental evidence is required to validate this hypothesis, such as site-directed mutagenesis of catalytic residues in addition to solving the structure of a CDTa–actin complex.

CONCLUDING REMARKS AND PERSPECTIVES

Some of the more dominant and pathogenic strains of *C. difficile* produce two large glucosylating toxins (Toxin A and Toxin B) and an ADPRT toxin (CDT) such as that from the CD196 strain [10]. However, there are strains of *C. difficile* that have variations in their ability to produce these toxins, and there is a great uncertainty over the roles of each of these toxins individually in pathogenesis [21–23]. As some of the most pathogenic strains of *C. difficile* produce CDT in addition to the LCTs, it is tempting to question whether or not CDT plays an adjunctive role to the LCTs in pathogenesis.

In the present review we have summarised the research into clostridial toxin structures. The full-length structures of the LCTs are yet to be determined, but the structures of some of the individual domains of these toxins have been solved. From the current structures we now have a better understanding of the binding, the autoproteolysis and their mechanism of action once the enzymatic domain has entered the cytosol. These findings have opened up a number of avenues that are of great value for use in drug therapeutics. These structures have provided multiple targets

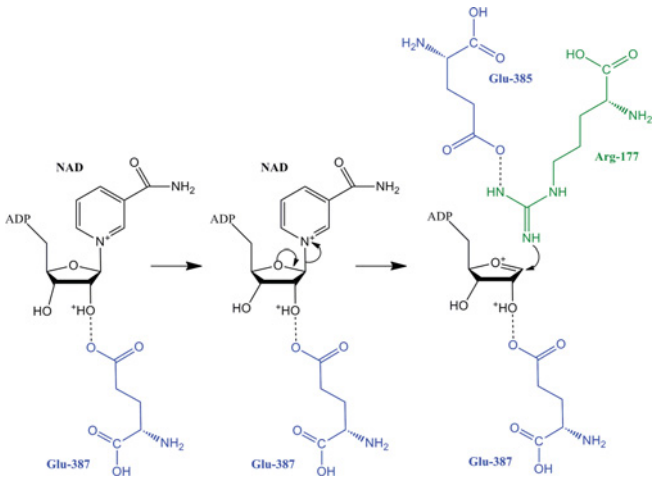


Figure 7 Predicted mechanism of ADP-ribosylation in CDT

The catalytic glutamate residue (blue) forms a hydrogen bond with the 2'-OH of the ribose group in NAD (black). This stabilizes the oxocarbenium intermediate leaving the glycosidic bond susceptible to nucleophilic attack from the target arginine residue from actin (green).

for drug design such as the binding domain to prevent uptake of the toxins, the cysteine protease domain to prevent cleavage of the toxin, or targeting the enzymatic domain to inhibit its activity. Determination of the crystal structures of the remaining unsolved domains is highly desirable, in addition to solving the structure of an enzyme–substrate complex. This will provide an understanding of how the toxins function as a whole. The crystal structure of CDT has provided a great insight into the active site of the toxin and its mechanistic implications. Further work using mutational analysis and enzyme–substrate complex formation will improve our understanding of the mechanism of ADP-ribosylation and help provide more specific drug targets for therapeutic design.

OUTSTANDING QUESTIONS

(1) What renders Toxin B more cytotoxic than Toxin A?

(2) How do the LCTs bind to cell surfaces, do their specificities differ and does this affect their differences in pathogenicity?

(3) How does the cysteine protease 'C' domain cleave the toxin and why does it require InsP_6?

(4) By what mechanism do the LCTs glucosylate Rho GTPases?

(5) What is the precise role of CDT in pathogenesis?

(6) Could CDT play an adjunctive role to the LCTs in pathogenesis?

(7) By what mechanism does CDT ADP-ribosylate actin?

FUNDING

A.H.D. is supported by a joint post-graduate studentship by the BBSRC (U.K.) and Health Protection Agency (UK).

REFERENCES

1 Kelly, C. P. and LaMont, J. T. (1998) *Clostridium difficile* infection. Annu. Rev. Med. **49**, 375–390

2 McMaster-Baxter, N. L. and Musher, D. M. (2007) *Clostridium difficile*: recent epidemiologic findings and advances in therapy. Pharmacotherapy **27**, 1029–1039

3 Ananthakrishnan, A. N. (2011) *Clostridium difficile* infection: epidemiology, risk factors and management. Nat. Rev. Gastroenterol. Hepatol. **8**, 17–26

4 Bartlett, J. G. (2002) Clinical practise. Antibiotic-associated diarrhea. New Engl. J. Med. **346**, 334–339

5 Lyerly, D. M., Krivan, H. C. and Wilkins, T. D. (1988) *Clostridium difficile*: its disease and toxins. Clin. Microbiol. Rev. **1**, 1–18

6 Voth, D. E. and Ballard, J. D. (2005) *Clostridium difficile* toxins: mechanism of action and role in disease. Clin. Microbiol. Rev. **18**, 247–263

7 Borriello, S. P. and Wilcox, M. H. (1998) *Clostridium difficile* infections of the gut: the unanswered questions. J. Antimicrob. Chemother. **41**, 67–69

8 Knoop, F. C., Owens, M. and Crocker, I. C. (1993) *Clostridium difficile*: clinical disease and diagnosis. Clin. Microbiol. Rev. **6**, 251–265

9 Borriello, S. P. (1998) Pathogenesis of *Clostridium difficile* infection. J. Antimicrob. Chemother. **41**, 13–19

10 Perelle, S., Gibert, M., Bourlioux, P., Corthier, G. and Popoff, M. R. (1997) Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* cd196. Infect. Immun. **65**, 1402–1407

11 Popoff, M. R., Rubin, E. J., Gill, D. M. and Boquet, P. (1988) Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. Infect. Immun. **56**, 2299–2306

12 Pellizzari, R., Rossetto, O., Schiavo, G. and Montecucco, C. (1999) Tetanus and botulinum neurotoxins: mechanism of action and therapeutic uses. Philos. Trans. R. Soc. Lond. B Biol. Sci. **354**, 259–268

13 Jank, T. and Aktories, K. (2008) Structure and mode of action of clostridial glucosylating toxins: the ABCD model. Trends Microbiol. **16**, 222–229

14 Busch, C. and Aktories, K. (2000) Microbial toxins and the glycosylation of Rho family GTPases. Curr. Opin. Struct. Biol. **10**, 528–535

15 Just, I. and Gerhard, R. (2004) Large clostridial cytotoxins. Rev. Physiol. Biochem. Pharmacol. **152**, 23–47

16 Rupnik, M., Dupuy, B., Fairweather, N. F., Gerding, D. N., Johnson, S., Just, I., Lyerly, D. M., Popoff, M. R., Rood, J. I., Sonenshein, A. L. et al. (2005) Revised nomenclature of *Clostridium difficile* toxins and associated genes. J. Med. Microbiol. **54**, 113–117

17 Tan, K. S., Wee, B. Y. and Song, K. P. (2001) Evidence for holin function of tcdE gene in the pathogenicity of *Clostridium difficile*. J. Med. Microbiol. **50**, 613–619

18 Rupnik, M., Avesani, V., Janc, M., von Eichel-Streiber, C. and Delmée, M. (1998) A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. J. Clin. Microbiol. **36**, 2240–2247

19 Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. and Minton, N. P. (2007) The ClosTron: a universal gene knock-out system for the genus *Clostridium*. J. Microbiol. Methods **70**, 452–464

20 O'Connor, J. R., Lyras, D., Farrow, K. A., Adams, V., Powell, D. R., Hinds, J., Cheung, J. K. and Rood, J. I. (2006) Construction and analysis of chromosomal *Clostridium difficile* mutants. Mol. Microbiol. **61**, 1335–1351

21 Lyerly, D. M., Saum, K. E., MacDonald, D. K. and Wilkins, T. D. (1985) Effects of *Clostridium difficile* toxins given intragastrically to animals. Infect. Immun. **47**, 349–352

22 Lyras, D., O'Connor, J. R., Howarth, P. M., Sambol, S. P., Carter, G. P., Phumoonna, T., Poon, R., Adams, V., Vedantam, G., Johnson, S. et al. (2009) Toxin B is essential for virulence of *Clostridium difficile*. Nature **458**, 1176–1179

23 Kuehne, S. A., Cartman, S. T., Heap, J. T., Kelly, M. L., Cockayne, A. and Minton, N. P. (2010) The role of Toxin A and Toxin B in *Clostridium difficile* infection. Nature **467**, 711–713

24 Frisch, C., Gerhard, R., Aktories, K., Hofmann, F. and Just, I. (2003) The complete receptor-binding domain of *Clostridium difficile* Toxin A is required for endocytosis. Biochem. Biophys. Res. Commun. **300**, 706–711

25 Barth, H., Pfeifer, G., Hofmann, F., Maier, E., Benz, R. and Aktories, K. (2001) Low pH-induced formation of ion channels by *Clostridium difficile* Toxin B in target cells. J. Biol. Chem. **276**, 10670–10676

26 Qa'Dan, M., Spyres, L. M. and Ballard, J. D. (2000) pH-induced conformational changes in *Clostridium difficile* Toxin B. Infect. Immun. **68**, 2470–2474

27 Egerer, M., Giesemann, T., Jank, T., Satchell, K. J. F. and Aktories, K. (2007) Auto-catalytic cleavage of *Clostridium difficile* Toxins A and B depends on cysteine protease activity. J. Biol. Chem. **282**, 25314–25321

28 Giesemann, T., Egerer, M., Jank, T. and Aktories, K. (2008) Processing of *Clostridium difficile* toxins. J. Med. Microbiol. **57**, 690–696

29 Pruitt, R. N., Chagot, B., Cover, M., Chazin, W. J., Spiller, B. and Lacy, D. B. (2009) Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in *Clostridium difficile* Toxin A. J. Biol. Chem. **284**, 21934–21940

30 Jank, T., Giesemann, T. and Aktories, K. (2007) Rho-glucosylating *Clostridium difficile* Toxins A and B: new insights into structure and function. Glycobiology **17**, 15R–22R

31 Just, I., Wilm, M., Selzer, J., Rex, G., von Eichel-Streiber, C., Mann, M. and Aktories, K. (1995) The enterotoxin from *Clostridium difficile* (ToxA) monoglucosylates the Rho proteins. J. Biol. Chem. **270**, 13932–13936

32 Just, I., Selzer, J., Wilm, M., von Eichel-Streiber, C., Mann, M. and Aktories, K. (1995) Glucosylation of Rho proteins by *Clostridium difficile* Toxin B. Nature **375**, 500–503

33 Pruitt, R. N., Chambers, M. G., Ng, K. K., Ohi, M. D. and Lacy, D. B. (2010) Structural organization of the functional domains of *Clostridium difficile* Toxins A and B. Proc. Natl. Acad. Sci. U.S.A. **107**, 13467–13472

34 Albesa-Jové, D., Bertrand, T., Carpenter, E. P., Swain, G. V., Lim, J., Zhang, J., Haire, L. F., Vasisht, N., Braun, V., Lange, A. et al. (2010) Four distinct structural domains in *Clostridium difficile* Toxin B visualized using SAXS. J. Mol. Biol. **396**, 1260–1270

35 Reinert, D. J., Jank, T., Aktories, K. and Schulz, G. E. (2005) Structural basis for the function of *Clostridium difficile* Toxin B. J. Mol. Biol. **351**, 973–981

36 Ciesla, W. P. J. and Bobak, D. A. (1998) *Clostridium difficile* Toxins A and B are cation-dependent UDP-glucose hydrolases with differing catalytic activities. J. Biol. Chem. **273**, 16021–16026

37 Rupnik, M., Pabst, S., Rupnik, M., von Eichel-Streiber, C., Urlaub, H. and Söling, H. (2005) Characterization of the cleavage site and function of resulting cleavage fragments after limited proteolysis of *Clostridium difficile* Toxin B (TcdB) by host cells. Microbiology **151**, 199–208

38 Unligil, U. M. and Rini, J. M. (2000) Glycosyltransferase structure and mechanism. Curr. Opin. Struct. Biol. **10**, 510–517

39 Jank, T., Reinert, D. J., Giesemann, T., Schulz, G. E. and Aktories, K. (2005) Change of the donor substrate specificity of *Clostridium difficile* Toxin B by site-directed mutagenesis. J. Biol. Chem. **280**, 37833–37838

40 Bishop, A. L. and Hall, A. (2000) Rho GTPases and their effector proteins. Biochem. J. **348**, 241–255

41 Jank, T., Giesemann, T. and Aktories, K. (2007) *Clostridium difficile* glucosyltransferase Toxin B-essential amino acids for substrate binding. J. Biol. Chem. **282**, 35222–35231

42 Krivan, H. C., Clark, G. F., Smith, D. F. and Wilkins, T. D. (1986) Cell surface binding site for *Clostridium difficile* enterotoxin: evidence for a glycoconjugate containing the sequence Galα1-3Galβ1–4GlcNAc. Infect. Immun. **53**, 573–581

43 Ho, J. G. S., Greco, A., Rupnik, M. and Ng, K. K. (2005) Crystal structure of receptor-binding C-terminal repeats from *Clostridium difficile* Toxin A. Proc. Natl. Acad. Sci. U.S.A. **102**, 18373–18378

44 Greco, A., Ho, J. G. S., Lin, S., Palcic, M. M., Rupnik, M. and Ng, K. K. (2006) Carbohydrate recognition by *Clostridium difficile* Toxin A. Nat. Struct. Mol. Biol. **13**, 460–461

45 Dove, C. H., Wang, S. Z., Price, S. B., Phelps, C. J., Lyerly, D. M., Wilkins, T. D. and Johnson, J. L. (1990) Molecular characterization of the *Clostridium difficile* Toxin A gene. Infect. Immun. **58**, 480–488

46 Kreimeyer, I., Euler, F., Marckscheffel, A., Tatge, H., Pich, A., Olling, A., Schwarz, J., Just, I. and Gerhard, R. (2011) Autoproteolytic cleavage mediates cytotoxicity of *Clostridium difficile* Toxin A. Naunyn Schmiedebergs Arch. Pharmacol. **383**, 253–262

46a Shen, A., Lupardus, P. J., Gersch, M. M., Puri, A. W., Albrow, V. E., Garcia, K. C. and Bogyo, M. (2011) Defining an allosteric circuit in the cysteine protease domain of *Clostridium difficile* Toxins. Nat. Struct. Mol. Biol. **18**, 364–371

47 Holbourn, K. P., Shone, C. C. and Acharya, K. R. (2006) A family of killer toxins. Exploring the mechanism of ADP-ribosylating toxins. FEBS J. **273**, 4579–4593

48 Barth, H. (2004) Uptake of binary actin ADP-ribosylating toxins. Rev. Physiol. Biochem. Pharmacol. **152**, 165–182

49 Geric, B., Johnson, S., Gerding, D. N., Grabnar, M. and Rupnik, M. (2003) Frequency of binary toxin genes among *Clostridium difficile* strains that do not produce large clostridial toxins. J. Clin. Microbiol. **41**, 5227–5232

50 Sundriyal, A., Roberts, A. K., Ling, R., McGlashan, J., Shone, C. C. and Acharya, K. R. (2010) Expression, purification and cell cytotoxicity of actin-modifying binary toxin from *Clostridium difficile*. Protein Expr. Purif. **74**, 42–48

51 Schwan, C., Stecher, B., Tzivelekidis, T., van Ham, M., Rohde, M., Hardt, W., Wehland, J. and Aktories, K. (2009) *Clostridium difficile* toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria. PLoS Pathog. **5**, e1000626

52 Barth, H., Aktories, K., Popoff, M. R. and Stiles, B. G. (2004) Binary bacterial toxins: biochemistry, biology, and applications of common *Clostridium* and *Bacillus* proteins. Microbiol. Mol. Biol. Rev. **68**, 373–402

53 Gülke, I., Pfeifer, G., Liese, J., Fritz, M., Hofmann, F., Aktories, K. and Barth, H. (2001) Characterization of the enzymatic component of the ADP-ribosyltransferase Toxin CDTa from *Clostridium difficile*. Infect. Immun. **69**, 6004–6011

54 Mauss, S., Chaponnier, C., Just, I., Aktories, K. and Gabbiani, G. (1990) ADP-ribosylation of actin isoforms by *Clostridium botulinum* C2 Toxin and *Clostridium perfringens* iota toxin. Eur. J. Biochem. **194**, 237–241

55 Vandekerckhove, J., Schering, B., Bärmann, M. and Aktories, K. (1987) *Clostridium perfringens* iota toxin ADP-ribosylates skeletal muscle actin in Arg-177. FEBS Lett. **225**, 48–52

56 Sundriyal, A., Roberts, A. K., Shone, C. C. and Acharya, K. R. (2009) Structural basis for substrate recognition in the enzymatic component of ADP-ribosyltransferase toxin CDTa from *Clostridium difficile*. J. Biol. Chem. **284**, 28713–28719

57 Tsuge, H., Nagahama, M., Oda, M., Iwamoto, S., Utsunomiya, H., Marquez, V. E., Katunuma, N., Nishizawa, M. and Sakurai, J. (2008) Structural basis of actin recognition and arginine ADP-ribosylation by *Clostridium perfringens* iota-toxin. Proc. Natl. Acad. Sci. U.S.A. **105**, 7399–7404

Received 18 January 2011/4 February 2011; accepted 21 February 2011
Published on the Internet 27 May 2011, doi:10.1042/BJ20110106